

STIC-Biotech/ChemLib

77769

Fr m: Ford, Vanessa
Sent: Thursday, October 10, 2002 5:38 PM
To: STIC-Biotech/ChemLib
Subject: In re: 09543407 Sequence search

Please search SEQ ID NO: 5 and please run an oligmer search on SEQ ID NO: 5.

I also need interference searches run.

Thanks!

Vanessa L. Ford
Biotechnology Patent Examiner
Office: CM1 8B13
Mailbox: CM1 8E12
Phone: 703.308.4735

RECEIVED
OCT 11 2002
STIC

Edward Hart
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 10/11/02
Date Completed: 10/15/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: 7 _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: PS _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: October 11, 2002, 21:14:28 ; Search time 33 Seconds
(without alignments)
508.247 Million cell updates/sec

Title: US-09-543-407-5
Perfect score: 151
Sequence: 1 MKLKVAAFAIVVSGSALA.....DSSVMVHQVGFNNATANQY 151

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size: 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: A_Geneseq_032802.*

```
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	151	100.0	151	15	AA874625
2	151	100.0	151	21	AAB36341
3	127	84.1	151	21	AAB36346
4	122	80.8	151	18	AAW23570
5	118	78.1	151	21	AAB36347
6	113	74.8	151	21	AAB36348
7	112	74.2	120	15	AA862761
8	112	74.2	120	18	AAW23569
9	110	72.6	151	21	AAB36352
10	102	67.5	151	21	AAB36355
11	94	62.3	151	21	AAB36349

12	87	57.6	151	21	AAB36351
13	80	53.0	151	21	AAB36354
14	78	51.7	151	21	AAB36353
15	70	46.4	151	21	AAB36350
16	45	29.8	151	21	AAB36316
17	23	15.2	23	21	AAB36319
18	23	15.2	23	21	AAB36321
19	23	15.2	23	21	AAB36324
20	23	15.2	23	21	AAB36326
21	23	15.2	23	21	AAB36338
22	23	15.2	23	21	AAB36340
23	22	14.6	22	21	AAB36318
24	22	14.6	22	21	AAB36320
25	22	14.6	22	21	AAB36322
26	22	14.6	22	21	AAB36325
27	22	14.6	22	21	AAB36327
28	22	14.6	22	21	AAB36337
29	22	14.6	22	21	AAB36339
30	19	12.6	19	21	AAB36323
31	19	12.6	19	21	AAB36328
32	19	12.6	19	21	AAB36336
33	11	7.3	142	15	AAR52664
34	11	7.3	151	21	AAB36343
35	10	6.6	19	21	AAB36333
36	9	6.0	22	21	AAB36330
37	9	6.0	122	15	AAR52663
38	7	4.6	64	21	AA802138
39	7	4.6	80	19	AAW38495
40	7	4.6	145	22	AAU01009
41	7	4.6	204	22	AA825515
42	7	4.6	206	22	AA891579
43	7	4.6	233	22	AA850247
44	7	4.6	257	20	AAW74001
45	7	4.6	257	20	AAV59838

ALIGNMENTS

```
RESULT 1
AAR74625 standard; Protein: 151 AA.
AAR74625:
AC AAR74625:
DT 26-JUN-1995 (first entry)
DE Agfa sequence.
KM Salmonella; Agfa; vaccine.
OS Salmonella.
XX
XX W09425598-A.
XX
XX 10-NOV-1994.
XX
XX 26-APR-1994; 94WO-IB00207.
XX
XX 26-APR-1994; 93US-0054452.
XX
XX 26-APR-1993; 93US-0054452.
XX
XX (KING/) KING J.
XX (UYVT-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
XX Clouthier SC, Collinson SK, Doran JL, Kay WM;
XX WPI; 1994-358275/44.
XX N-PSDB; AA087467.
XX
XX Eliciting an immune response to Salmonella - using attenuated
XX Salmonella strains, vector constructs, or compsns. contg.
XX fimbrial type proteins.
XX
```

PS Disclosure: Fig7B: 95pp; English.

XX The Salmonella Agfa protein and DNA are used in vaccine and

CC genetic immunization compositions, respectively, to elicit an

CC immune response to salmonella in animals (e.g. food producing

CC animals) and humans.

XX

SQ Sequence 151 AA;

Query Match 100.0%; Score 151; DB 15; Length 151;

Best Local Similarity 100.0%; Pred. No. 3e-144;

Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLKVAFAFAIVVSGSALAGVVPOMGGGNGHNGGNSGSPDSTLSIYOGSANAALALQ 60

DB 1 MKLKVAFAFAIVVSGSALAGVVPOMGGGNGHNGGNSGSPDSTLSIYOGSANAALALQ 60

QY 61 SPARKSETTITQSGYNGADVGGADNSTIELTONGFRNNATTIDQWNAKNSDITVGOYGG 120

DB 61 SPARKSETTITQSGYNGADVGGADNSTIELTONGFRNNATTIDQWNAKNSDITVGOYGG 120

QY 121 NNAALVNOTASDSSVMVROYGFGNNATANOX 151

DB 121 NNAALVNOTASDSSVMVROYGFGNNATANOX 151

RESULT 2

AAB36341 standard; Protein: 151 AA.

XX AAB36341;

XX 26-FEB-2001 (first entry)

DE Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.

XX

DE Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;

KW vaccine; immune response; Immunogen.

KW

XX Salmonella enteritidis.

OS

XX WO200060102-A2.

PN

XX 12-OCT-2000.

PD

XX 05-APR-2000; 2000WO-CA00356.

PF

XX 05-APR-1999; 99US-0127888.

PR

XX (UYVI-) UNIV VICTORIA.

PA

XX White AP, Doran JL, Collison SK, Kay WW;

PI

XX WPI: 2000-672631/65.

DR

XX N-PSDB: AAC64617.

PT

XX Recombinant agfa gene having a segment replaced by a foreign DNA

PT sequence which encodes foreign epitope or antigen, expresses

PT recombinant Agfa protein useful for eliciting immune response in animal

PT

XX

PS Disclosure: Page 135; 139pp; English.

XX

XX The present invention describes a recombinant agfa gene (1) where a

CC segment of the gene has been replaced by a segment of a foreign DNA

CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF1/7AF) nucleation depended

CC assembly system of strains of Salmonella, Escherichia coli and

CC Enterobacteriaceae for the production of fimbriae comprising recombinant

CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene

CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,

CC comprising separating an amino acid polymer comprising a recombinant

CC Agfa protein containing a replacement segment or segments of foreign

CC amino acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the

CC polymer into the animal in conjunction with a carrier or diluent. (1) is

CC useful for the expression of recombinant Agfa protein which is useful

CC for eliciting an immune response in an animal. In a fimbrial presentation

CC system the heterologous antigens are presented in high numbers (up to

CC 500,000 copies/cell), the hybrid fimbria protein possesses both the

CC immunogenicity and adhesion properties relevant for an efficient live

CC vaccine, the carrier fimbrial subunit proteins are usually strong

CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fimbriae are easy and

CC inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention.

XX

SQ Sequence 151 AA;

Query Match 100.0%; Score 151; DB 21; Length 151;

Best Local Similarity 100.0%; Pred. No. 3e-144;

Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLKVAFAFAIVVSGSALAGVVPOMGGGNGHNGGNSGSPDSTLSIYOGSANAALALQ 60

DB 1 MKLKVAFAFAIVVSGSALAGVVPOMGGGNGHNGGNSGSPDSTLSIYOGSANAALALQ 60

QY 61 SPARKSETTITQSGYNGADVGGADNSTIELTONGFRNNATTIDQWNAKNSDITVGOYGG 120

DB 61 SPARKSETTITQSGYNGADVGGADNSTIELTONGFRNNATTIDQWNAKNSDITVGOYGG 120

QY 121 NNAALVNOTASDSSVMVROYGFGNNATANOX 151

DB 121 NNAALVNOTASDSSVMVROYGFGNNATANOX 151

RESULT 3

AAB36346 standard; Protein: 151 AA.

XX AAB36346;

XX 26-FEB-2001 (first entry)

DE Agfa::PT3#1 amino acid sequence SEQ ID NO:12.

XX

DE Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;

KW vaccine; immune response; Immunogen.

KW

XX Salmonella enteritidis.

OS

XX Escherichia coli.

OS

XX Synthetic.

XX

XX WO200060102-A2.

PN

XX 12-OCT-2000.

PD

XX 05-APR-2000; 2000WO-CA00356.

PF

XX 05-APR-1999; 99US-0127888.

PR

XX (UYVI-) UNIV VICTORIA.

PA

XX White AP, Doran JL, Collison SK, Kay WW;

PI

XX WPI: 2000-672631/65.

DR

XX N-PSDB: AAC64622.

PT

XX Recombinant agfa gene having a segment replaced by a foreign DNA

PT sequence which encodes foreign epitope or antigen, expresses

PT recombinant Agfa protein useful for eliciting immune response in animal

PT

XX

PS Disclosure: Page 135; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF1/77A) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant
CC AgfA protein containing a replacement segment or segments of foreign
CC amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful
CC for eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention.

XX Sequence 151 AA:

Query Match 84.1%; Score 127; DB 21; Length 151;

Best Local Similarity 100.0%; Pred. No. 4; Be-120; Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKLKVAAFAAIVSGSALAGVPPOMGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
DB 1 MKLKVAAFAAIVSGSALAGVPPOMGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
OY 61 SPARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQWKNKNSDITVGQYCG 120
DB 61 SPARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQWKNKNSDITVGQYCG 120
OY 121 NNALVN 127
DB 121 NNALVN 127

XX

RESULT 4
ID AAW23570 standard; Protein: 151 AA.

XX AAW23570:

XX 29-SEP-1997 (first entry)

DE *Salmonella enteritidis* 27655-3b agfA.

KM Enteropathogenic bacteria; enterobacteria; *S. enteritidis*; antibody.

OS *Salmonella enteritidis*.

XX Key Location/Qualifiers

FT Misc-difference 123 /note= "Encoded by GCC"

XX US5635617-A.

XX 03-JUN-1997.

XX 26-APR-1993; 93US-0054452.

XX 26-APR-1994; 94US-0233788.

PR 26-APR-1993; 93US-0054452.
XX
XX (UUVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA Clouthier SC, Collinson SK, Doran JL, Kay WM;
XX
XX WPI: 1997-309866/28.
DR N-PSDB; AAT74142.
XX
XX Isolated *Salmonella* gene agfA - used for diagnosis of *Salmonella* or
XX enteropathogenic bacteria of the Enterobacteria family
XX Example 2; Fig 7; 85pp; English.

XX The present sequence represents agfA encoded by the full agfA gene
XX derived from *Salmonella enteritidis* 27655-3b. The nucleic acid can
XX be used to provide diagnostic assays for *Salmonella* and/or
XX enteropathogenic bacteria of the family Enterobacteria. It can also
XX be used to provide proteins and antibodies which can be used for
XX assays. The nucleic acid sequence can be used to provide probes or
XX primers which can specifically hybridise to nucleic acid molecules
XX from greater than 98% of *Salmonella* strains that are pathogenic to
XX warm-blooded animals relative to nucleic acid molecules from
XX virtually all other microbial organisms.

XX Sequence 151 AA:

Query Match 80.8%; Score 122; DB 18; Length 151;

Best Local Similarity 100.0%; Pred. No. 5; Be-115; Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKLKVAAFAAIVSGSALAGVPPOMGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
DB 1 MKLKVAAFAAIVSGSALAGVPPOMGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
OY 61 SPARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQWKNKNSDITVGQYCG 120
DB 61 SPARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQWKNKNSDITVGQYCG 120
OY 121 NN 122
DB 121 NN 122

XX

RESULT 5
ID AAB36347 standard; Protein: 151 AA.

XX AAB36347:

XX 26-FEB-2001 (first entry)

DE AgfA::PT3#2 amino acid sequence SEQ ID NO:14.

KM *Salmonella*; agfA; chromosomal gene replacement; fimbria; epitope;

KW vaccine; immune response; immunogen.

OS *Salmonella enteritidis*.

OS *Escherichia coli*.

XX Synthetic.

XX WO200060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA00356.

XX 05-APR-1999; 99US-0127888.

XX (UUVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collinson SK, Kay WM;
XX

DR	WP1: 2000-672631/65.
NR	N-PSDB; AAC64623.
XX	Recombinant agfa gene having a segment replaced by a foreign DNA
PT	Sequence which encodes foreign epitope or antigen, expresses
PT	recombinant Agfa protein useful for eliciting immune response in animal
PS	-
XX	Disclosure; Page 136; 139pp; English.
XX	
CC	The present invention describes a recombinant agfa gene (I) where a
CC	segment of the gene has been replaced by a segment of a foreign DNA
CC	sequence which encodes a foreign epitope or antigen. Also described are:
CC	(1) use of thin aggregative fimbriae (SEF1/7AP) nucleation depended
CC	assembly system of strains of Salmomella, Escherichia coli and
CC	Enterobacteriaceae for the production of fimbriae comprising recombinant
CC	AgfA, CsgA and Agfa-homologue fibrin subunits, respectively; (2)
CC	directing recombination of a recombinant gene into the chromosome of the
CC	homologous species; (3) directing recombination of a recombinant gene
CC	back into the chromosome of the homologous species, replacing the native
CC	copy of that gene; and (4) eliciting an immune response in an animal,
CC	comprising separating an amino acid polymer comprising a recombinant
CC	Agfa protein containing a replacement segment or segments of foreign
CC	amino acid sequence or sequences grown on a Salmomella, E. coli or
CC	Enterobacteriaceae host cell, from the host cell and introducing the
CC	polymer into the animal in conjunction with a carrier or diluent. (I) is
CC	useful for the expression of recombinant Agfa protein which is useful
CC	for eliciting an immune response in an animal. In a fibrillar presentation
CC	system the heterologous antigens are presented in high numbers (up to
CC	500,000 copies/cell), the hybrid fibrin protein possesses both the
CC	immunogenicity and adhesion properties relevant for an efficient live
CC	vaccine, which may be important for directing an immune response
CC	against the inserted epitope, and hybrid fimbriae are easy and
CC	inexpensive to purify in large amount. The present sequence is given in
CC	the exemplification of the present invention.
XX	
SQ	Sequence 151 AA:
OY	Query Match 78.1%; Score 118; DB 21; Length 151;
	Best Local Similarity 100.0%; Pred. No. 5.7e-111; Mismatches 118; Conservative 0; Indels 0; Gaps 0
DG	1 MKILKVAFAPIVVGSGALAGVVPQMGGGNGHNGSGSPDSTLTIYXGSANALALAQ 60
OY	1 MKILKVAFAPIVVGSGALAGVPGMGGGNGHNGSGSPDSTLTIYXGSANALALAQ 60
DG	1 MKELKVAFAPIVVGSGALAGVPMGGGNGHNGSGSPDSTLTIYXGSANALALAQ 60
OY	61 SDARKSETTIGSYGNGADYGCGADNSTIFLLONGFRNNATIDQWAKNSDTIVGOY 118
DG	61 SDARKSETTIGSYGNGADYGCGADNSTIFLLONGFRNNATIDQWAKNSDTIVGOY 118
RESULT 6	
AAB36348	ID AAB36348 standard; Protein: 151 AA.
XX	
AC	AAB36348;
DT	26-FEB-2001 (first entry)
XX	
DE	AgfA::PT3#3 amino acid sequence SEQ ID NO:16.
XX	
KW	Salmomella; agfa; chromosomal gene replacement; fibrin; epitope;
KW	vaccine; immune response; immunogen.
XX	
OS	Salmomella enteritidis.
OS	Escherichia coli.
OS	Synthetic.
XX	
PN	WO200060102-A2.
XX	
PD	12-OCT-2000.
JM	

```

05-APR-2000: 2000MO-CA00356.
XX
PR -05-APR-1999: 99US-0127888.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay MW;
XX
DR WPI: 2000-672631/65.
XX
N-PSDB: AAC64624.
XX
Recombinant agfa gene having a segment replaced by a foreign DNA
PT sequence which encodes foreign epitope or antigen, expresses
recombinant Agfa protein useful for eliciting immune response in animal
PT
-
Disclosure: Page 136; 139pp; English.
XX
PS
XX
The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC sequence (1) use of thin aggregative fimbriae (SEF1/rAF) nucleation depended
CC (1) use of thin aggregative fimbriae (SEF1/rAF) nucleation depended
CC assembly system of strains of Salmomella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC fimbrial subunits, respectively; (2)
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating a replacement segment or segments of foreign
CC Agfa protein containing a replacement segment or segments of foreign
CC amino acid sequence or sequences grown on a Salmomella, E. coli or
CC enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful
CC for eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention.
XX
SQ Sequence 151 AA:
XX
Query Match 74.8%; Score 113; DB 21; Length 151;
Best Local Similarity 100.0%; Pred. No. 6.3e-106; Indels 0; Gaps 0;
Matches 113; Conservative 0; Mismatches 0;
0Y 39 SGPSTLSITYOXSANMALALOSDARKSETTITQSGGNGCADVGOGADNSTIELTONGFR 98
Db |-----|
39 SGPSTLSITYOXSANMALALQSDARKSETTITQSGGNCADVGGAGDNSTIELTONGFR 98
0Y NNATIDOMNAKNSDITVGOYGNNALVNQTASDVSSVMVRQVGFGNATTANOY 151
Db |||||||||
99 NNATIDOMNAKNSDITVGOYGNNALVNQTASDVSSVMVRQVGFGNATTANOY 151
Db |||||||||
RESULT 7
ID AAR62761 standard; Protein: 120 AA.
AC AAR62761;
XX
DT 26-JUN-1995 (first entry)
XX
DE Agfa sequence.
XX
KW Salmomella; Agfa; vaccine.
XX
OS Salmomella enteritis 27655-3b.

```

XX WO9425598-A.
PN 10-NOV-1994.
XX
XX 26-APR-1994; 94AO-IB00207.
XX
XX 26-APR-1994; 93US-0054452.
XX
XX (KING/) KING J.
PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI Clouthier SC, Collinson SK, Doran JL, Kay WW;
XX WPI: 1994-358275/44.
DR N-PSDB: AAO73066.
XX
PT Eliciting an immune response to *Salmonella* - using attenuated
PT *Salmonella* strains, vector constructs, or comps. contg.
PT fimbrial type proteins.
PS
XX Disclosure: Fig7A: 95pp; English.
XX
CC The sequence represents the *Salmonella enteritis* 27655-3b
CC *TnphoA* mutant strain Agfa protein. The encoding DNA and
CC isolated Agfa protein are used in genetic immunization and vaccine
CC compositions, respectively, to elicit an immune response to
CC *Salmonella* in animals (e.g. food producing animals) and humans.
XX
SQ Sequence 120 AA:

Query Match 74.2%; Score 112; DB 15; Length 120;
Best Local Similarity 100.0%; Pred. No. 5.2e-105;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 VVPOMGGGNNHNGGNSGPDSTLSIYQGSANALALQSDARKSETTTTQSGYNGADV 81
DB 1 VVPOMGGGNNHNGGNSGPDSTLSIYQGSANALALQSDARKSETTTTQSGYNGADV 60
OY 82 GOGADNSTIELTONGFRNATTIDOWNAKNSDITVGOYGNNALVNOTASDS 133
DB 61 GOGADNSTIELTONGFRNATTIDOWNAKNSDITVGOYGNNALVNOTASDS 112

RESULT 8

AAW23569
ID AAW23569 standard; Protein: 120 AA.

XX AAW23569;

DT 29-SEP-1997 (first entry)

XX *Salmonella enteritidis* 27655-3b *TnphoA* mutant agfa fragment.

XX Enteropathogenic bacteria; enterobacteria; *S. enteritidis*; antibody.

OS *Salmonella enteritidis*.

XX US5635617-A.

XX 03-JUN-1997.

XX 26-APR-1993; 93US-0054452.

XX 26-APR-1994; 94US-0233788.

XX 26-APR-1993; 93US-0034452.

XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.

XX Clouthier SC, Collinson SK, Doran JL, Kay WW;

XX WPI: 1997-309886/28.
XX N-PSDB: AAT74141.

XX Isolated *Salmonella* gene agfa - used for diagnosis of *Salmonella* or
PT enteropathogenic bacteria of the Enterobacteria family
XX
XX Example 2; Fig 7: 85pp; English.
XX
CC The present sequence represents an agfa fragment encoded by an agfa gene
CC fragment derived from *Salmonella enteritidis* 27655-3b *TnphoA* mutant
CC strain. The nucleic acid can be used to provide diagnostic assays for
CC *Salmonella* and/or enteropathogenic bacteria of the family
CC Enterobacteria. It can also be used to provide proteins and antibodies
CC which can be used for assays. The nucleic acid sequence can be used to
CC provide probes or primers which can specifically hybridise to nucleic
CC acid molecules from greater than 99% of *Salmonella* strains that are
CC pathogenic to warm-blooded animals relative to nucleic acid molecules
CC from virtually all other microbial organisms.
XX
SQ Sequence 120 AA:

Query Match 74.2%; Score 112; DB 18; Length 120;
Best Local Similarity 100.0%; Pred. No. 5.2e-105;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 VVPOMGGGNNHNGGNSGPDSTLSIYQGSANALALQSDARKSETTTTQSGYNGADV 81
DB 1 VVPOMGGGNNHNGGNSGPDSTLSIYQGSANALALQSDARKSETTTTQSGYNGADV 60
OY 82 GOGADNSTIELTONGFRNATTIDOWNAKNSDITVGOYGNNALVNOTASDS 133
DB 61 GOGADNSTIELTONGFRNATTIDOWNAKNSDITVGOYGNNALVNOTASDS 112

RESULT 9

AAB36352
ID AAB36352 standard; Protein: 151 AA.

XX AAB36352;

DT 26-FEB-2001 (first entry)

XX Agfa: PT3#7 amino acid sequence SEQ ID NO:24.

XX *Salmonella*: agfa; chromosomal gene replacement; fimbria; epitope;

XX vaccine; immune response; immunogen.

OS *Salmonella enteritidis*.

XX *Escherichia coli*.

XX Synthetic.

XX WO200060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA00356.

XX 05-APR-1999; 99US-0127888.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collinson SK, Kay WW;

XX WPI: 2000-672631/65.

XX N-PSDB: AAC64628.

XX Recombinant agfa gene having a segment replaced by a foreign DNA

XX sequence which encodes foreign epitope or antigen, expresses

XX recombinant Agfa protein useful for eliciting immune response in animal

XX Disclosure: Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a

XX segment of the gene has been replaced by a segment of a foreign DNA

sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant
 CC Agfa protein containing a replacement segment or segments of foreign
 CC amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful
 CC for eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention.

CC Sequence 151 AA:

Query Match 72.8%; Score 110; DB 21; Length 151;
 Best Local Similarity 100.0%; Pred. No. 6.7e-103; Indels 0; Gaps 0;
 Matches 110; Conservative 0; Mismatches 0;

OY 1 MKLLKVAAPFAIYVSSALAGVVPOMGGGNNHNGSGSPDSTLSIYOGSANAALALQ 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1 MKLLKVAAPFAIYVSSALAGVVPOMGGGNNHNGSGSPDSTLSIYOGSANAALALQ 60
 OY 61 SDRKSETTTQSGYGADVGAGADNSTIELTONGFRNNATIDQNAKN 110
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 61 SDRKSETTTQSGYGADVGAGADNSTIELTONGFRNNATIDQNAKN 110

RESULT 10

AAB36355 standard; Protein; 151 AA.

AC AAB36355;

DT 26-FEB-2001 (first entry)

XX Agfa::PT3#10 amino acid sequence SEQ ID NO:30.

XX *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;

KW vaccine; immune response; immunogen.

XX *Salmonella enteritidis*.

OS *Escherichia coli*.

OS Synthetic.

XX WO200060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA00356.

XX 05-APR-1999; 99US-0127888.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;

XX WPT. 2000-672631/65.

XX DR N-PSDB; AAC64631.

XX Recombinant agfa gene having a segment replaced by a foreign DNA

PT sequence which encodes foreign epitope or antigen, expresses
 PT recombinant Agfa protein useful for eliciting immune response in animal
 PS Disclosure: Page 139; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant
 CC Agfa protein containing a replacement segment or segments of foreign
 CC amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful
 CC for eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention.

XX Sequence 151 AA:

Query Match 67.5%; Score 102; DB 21; Length 151;
 Best Local Similarity 100.0%; Pred. No. 7.8e-95; Indels 0; Gaps 0;
 Matches 102; Conservative 0; Mismatches 0;

OY 1 MKLLKVAAPFAIYVSSALAGVVPOMGGGNNHNGSGSPDSTLSIYOGSANAALALQ 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1 MKLLKVAAPFAIYVSSALAGVVPOMGGGNNHNGSGSPDSTLSIYOGSANAALALQ 60
 OY 61 SDRKSETTTQSGYGADVGAGADNSTIELTONGFRNNAT 102
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 61 SDRKSETTTQSGYGADVGAGADNSTIELTONGFRNNAT 102

RESULT 11

AAB36349 standard; Protein; 151 AA.

AC AAB36349;

DT 26-FEB-2001 (first entry)

XX Agfa::PT3#4 amino acid sequence SEQ ID NO:18.

XX *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;

KW vaccine; immune response; immunogen.

XX *Salmonella enteritidis*.

OS *Escherichia coli*.

OS Synthetic.

XX WO200060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA00356.

XX 05-APR-1999; 99US-0127888.

PA (UYVI-) UNIV VICTORIA.
 PI White AP, Doran JL, Collison SK, Kay WM;
 XX
 XX MPI: 2000-672631/65.
 DR N-PSDB: AAC64625.
 XX
 XX Recombinant agfa gene having a segment replaced by a foreign DNA
 PT sequence which encodes foreign epitope or antigen, expresses
 PT recombinant agfa protein useful for eliciting immune response in animal
 PT
 PS Disclosure: Page 136; 139pp; English.
 XX
 XX The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant
 CC Agfa protein containing a replacement segment or segments of foreign
 CC amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful
 CC for eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention.
 XX
 SQ Sequence 151 AA:
 Query Match 62.3%; Score 94; DB 21; Length 151;
 Best Local Similarity 100.0%; Pred. No. 9.1e-87;
 Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 58 ALQSDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQNNAKNSDITVGO 117
 DB 58 ALQSDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQNNAKNSDITVGO 117
 QY 118 YCGNNAALVNOTASDSVYVROVGFGNNAATNOY 151
 DB 118 YCGNNAALVNOTASDSVYVROVGFGNNAATNOY 151
 RESULT 12
 AAB36351
 ID AAB36351 standard; Protein: 151 AA.
 XX
 XX AAB36351;
 AC
 XX
 XX 26-FEB-2001 (first entry)
 DT
 XX
 XX Agfa: PT3#6 amino acid sequence SEQ ID NO:22.
 DE
 XX
 XX *Salmonella*; agfa: chromosomal gene replacement; fimbrial; epitope;
 KW vaccine; immune response; immunogen.
 XX
 XX *Salmonella enteritidis*.
 OS
 XX *Escherichia coli*.
 OS
 XX Synthetic.

PN WO200060102-A2.
 XX
 XX 12-OCT-2000.
 PD
 XX
 XX 05-APR-2000; 2000WO-CA00356.
 PF
 XX
 XX 05-APR-1999; 99US-0127888.
 PR
 XX
 XX (UYVI-) UNIV VICTORIA.
 PA
 XX
 XX White AP, Doran JL, Collison SK, Kay WM;
 PI
 XX
 XX MPI: 2000-672631/65.
 DR
 XX
 XX N-PSDB: AAC64627.
 DR
 XX
 XX Recombinant agfa gene having a segment replaced by a foreign DNA
 PT sequence which encodes foreign epitope or antigen, expresses
 PT recombinant agfa protein useful for eliciting immune response in animal
 PT
 PS Disclosure: Page 137; 139pp; English.
 XX
 XX The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant
 CC Agfa protein containing a replacement segment or segments of foreign
 CC amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful
 CC for eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention.
 XX
 SQ Sequence 151 AA:
 Query Match 57.6%; Score 87; DB 21; Length 151;
 Best Local Similarity 100.0%; Pred. No. 1e-79;
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLKYAFAFAIYVSGSALAGVYVPOWGGGNNHGGNSSGPPSTLSIYQGSANALALQ 60
 DB 1 MKLKYAFAFAIYVSGSALAGVYVPOWGGGNNHGGNSSGPPSTLSIYQGSANALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGGADN 87
 DB 61 SDARKSETTITQSGYNGADVGGADN 87
 RESULT 13
 AAB36354
 ID AAB36354 standard; Protein: 151 AA.
 XX
 XX AAB36354;
 AC
 XX
 XX 26-FEB-2001 (first entry)
 DT
 XX
 XX Agfa: PT3#9 amino acid sequence SEQ ID NO:28.

XX	Salmonella: agfa: chromosomal gene replacement; fimbria; epitope:
KM	vaccine; immune response; immunogen.
XX	
OS	Salmonella enteritidis.
OS	Escherichia coli.
OS	Synthetic.
XX	
PN	WO20060102-A2.
XX	
PD	12-OCT-2000.
XX	
PF	05-APR-2000; 2000WO-CA00356.
XX	
PR	05-APR-1999; 99US-0127888.
XX	
PA	(UYVI-) UNIV VICTORIA.
PI	
XX	White AP, Doran JL, Collison SK, Kay WW:
DR	WPI: 2000-672631/65.
XX	N-PSDB; AAC64630.
XX	
PT	Recombinant agfa gene having a segment replaced by a foreign DNA
PT	sequence which encodes foreign epitope or antigen, expresses
PT	recombinant Agfa protein useful for eliciting immune response in animal
PS	-
XX	Disclosure: Page 138; 139pp; English.
XX	
CC	The present invention describes a recombinant agfa gene (I) where a
CC	segment of the gene has been replaced by a segment of a foreign DNA
CC	sequence which encodes a foreign epitope or antigen. Also described are:
CC	(1) use of thin aggregative fimbriae (SEF1/7A) nucleation depended
CC	assembly system of strains of Salmonella, Escherichia coli and
CC	Enterobacteriaceae for the production of fimbriae comprising recombinant
CC	Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC	directing recombination of a recombinant gene into the chromosome of the
CC	homologous species; (3) directing recombination of a recombinant gene
CC	back into the chromosome of the homologous species, replacing the native
CC	copy of that gene; and (4) eliciting an immune response in an animal,
CC	comprising separating an amino acid polymer comprising a recombinant
CC	Agfa protein containing a replacement segment or segments of foreign
CC	amino acid sequence or sequences grown on a Salmonella, E. coli or
CC	Enterobacteriaceae host cell, from the host cell and introducing the
CC	polymer into the animal in conjunction with a carrier or diluent. (I) is
CC	useful for the expression of recombinant Agfa protein which is useful
CC	for eliciting an immune response in an animal. In a fimbrial presentation
CC	system the heterologous antigens are presented in high numbers (up to
CC	500,000 copies/cell), the hybrid fimbria protein possesses both the
CC	immunogenicity and adhesion properties relevant for an efficient live
CC	vaccine, the carrier fimbrial subunit proteins are usually strong
CC	immunogens, which may be important for directing an immune response
CC	against the inserted epitope, and hybrid fimbriae are easy and
CC	inexpensive to purify in large amount. The present sequence is given in
CC	the exemplification of the present invention.
XX	
SQ	Sequence 151 AA:
XX	
Query Match	53.0%; Score 80; DB 21; Length 151;
Best Local Similarity	100.0%; Pred. No. 1.2e-72;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 MKLILVAAPAAALIVGSGSALAGVPPQMGGGGHNHGGNSGDPSTLSIYXGSANAALALQ 60
DB	1 MKLILVAAPAAALIVGSGSALAGVPPQMGGGGHNHGGNSGDPSTLSIYXGSANAALALQ 60
OY	61 SPARKSETTITGSGYNGAD 80
DB	61 SPARKSETTITGSGYNGAD 80

RESULT 14.

ID	Sequence	Score	DB	Length	Mismatches	Indels	Gaps
01	Query Match	51.7%	Score 78;	DB 21;	Length 151;		
02	Best Local Similarity	100.0%	Pred. No. 1.2e-70;				
03	Matches	78;	Conservative	0;	Mismatches	0;	Indels 0; Gaps 0
04	74	GGNGAGVGGAGDNGSTIELTONGFNNAATIDDMAKNSDRIIVGGGNNALVNGTASDS	133				

DB 74 GYGNGADVGQGDNSTIELTQNGFRNNATIDQWNAKNSDITVGYGGNNALVNOTASDS 133

QY 134 SYVVRQVGFGNNTANQY 151
 |||
 DB 134 SYVVRQVGFGNNTANQY 151

RESULT 15
 AAB36350
 ID AAB36350 standard; Protein: 151 AA.
 XX AAB36350;
 AC
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa: PT3#5 amino acid sequence SEQ ID NO:20.
 XX
 KM Salmoneila; agfa: chromosomal gene replacement; fimbria; epitope;
 KM vaccine; immune response; Immunogen.
 XX
 OS Salmoneila enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 FM WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA00356.
 XX
 PR 05-APR-1999; 99US-0127888.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 DR WPI: 2000-672631/65.
 DR N-PSDB: AAC64626.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA
 PT sequence which encodes foreign epitope or antigen, expresses
 PT recombinant agfa protein useful for eliciting immune response in animal
 PT
 PS
 XX
 PS Disclosure: Page 137; 139pp; English.
 CC
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmoneila, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant
 CC Agfa protein containing a replacement segment or segments of foreign
 CC amino acid sequence or sequences grown on a Salmoneila, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful
 CC for eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention.
 CC

SQ Sequence 151 AA:

Query Match 46.4%; Score 70; DB 21; Length 151;

Best Local Similarity 100.0%; Pred. No. 1.5e-62; Mismatches 0; Indels 0; Gaps 0;

Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 GGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGGNNALVNOTASDSVVRQV 141
 |||
 DB 82 GGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGGNNALVNOTASDSVVRQV 141
 QY 142 FGNNATANQY 151
 |||
 DB 142 FGNNATANQY 151

Search completed: October 11, 2002, 21:19:04
 Job time : 34 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 11, 2002, 21:18:23 ; Search time 17 Seconds
(without alignments)
216,957 Million cell updates/sec

Title: US-09-543-407-5

Perfect score: 151
Sequence: 1 MKLLKVAFAIVVSGSALA.....DSSVMRVGFGNNATANQY 151

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size: 0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/PCtUS.COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	122	80.8	151	1 US-08-233-788A-59	Sequence 59, Appl
2	112	74.2	120	1 US-08-233-788A-57	Sequence 57, Appl
3	7	4.6	80	1 US-08-858-207A-291	Sequence 291, App
4	7	4.6	304	1 US-07-851-976B-8	Sequence 8, Appli
5	7	4.6	304	1 US-08-291-609-8	Sequence 8, Appli
6	7	4.6	304	1 US-08-401-136-8	Sequence 8, Appli
7	7	4.6	304	3 US-08-850-554-8	Sequence 8, Appli
8	7	4.6	407	2 US-08-934-846-4	Sequence 4, Appli
9	7	4.6	407	2 US-09-238-557-4	Sequence 2, Appli
10	7	4.6	418	2 US-08-934-846-2	Sequence 2, Appli
11	7	4.6	418	2 US-09-238-557-2	Sequence 2, Appli
12	6	4.0	29	1 US-08-190-802A-97	Sequence 97, Appl
13	6	4.0	29	4 US-08-477-346-97	Sequence 97, Appl
14	6	4.0	29	4 US-08-473-089-97	Sequence 97, Appl
15	6	4.0	32	3 US-09-053-197A-61	Sequence 61, Appl
16	6	4.0	32	4 US-09-085-761A-66	Sequence 66, Appl
17	6	4.0	33	2 US-08-461-990B-10	Sequence 10, Appl
18	6	4.0	63	4 US-08-988-856B-29	Sequence 29, Appl
19	6	4.0	134	2 US-08-910-856-6	Sequence 67, Appl
20	6	4.0	157	4 US-09-025-769B-372	Sequence 372, App
21	6	4.0	157	4 US-09-025-769B-373	Sequence 373, App
22	6	4.0	159	3 US-09-027-449-53	Sequence 53, Appl
23	6	4.0	159	3 US-08-804-444A-53	Sequence 53, Appl
24	6	4.0	159	3 US-09-026-985-53	Sequence 53, Appl
25	6	4.0	163	4 US-09-025-769B-282	Sequence 282, App
26	6	4.0	164	4 US-08-484-993B-45	Sequence 45, Appl
27	6	4.0	164	2 US-08-484-158B-45	Sequence 45, Appl

28.	6	4.0	164	2 US-08-484-596A-45	Sequence 45, Appl
29	6	4.0	164	2 US-08-480-150A-45	Sequence 45, Appl
30	6	4.0	164	3 US-08-458-731-45	Sequence 45, Appl
31	6	4.0	164	3 US-08-149-223A-45	Sequence 45, Appl
32	6	4.0	188	2 US-08-484-993B-20	Sequence 20, Appl
33	6	4.0	188	2 US-08-484-158B-20	Sequence 20, Appl
34	6	4.0	188	2 US-08-484-596A-20	Sequence 20, Appl
35	6	4.0	188	2 US-08-480-150A-20	Sequence 20, Appl
36	6	4.0	188	3 US-08-458-731-20	Sequence 20, Appl
37	6	4.0	188	3 US-08-149-223A-20	Sequence 20, Appl
38	6	4.0	211	1 US-08-276-852-34	Sequence 34, Appl
39	6	4.0	211	1 US-08-133-011-16	Sequence 16, Appl
40	6	4.0	211	1 US-08-322-730A-16	Sequence 16, Appl
41	6	4.0	211	1 US-08-387-874-16	Sequence 16, Appl
42	6	4.0	211	1 US-08-899-575-34	Sequence 34, Appl
43	6	4.0	211	1 US-08-899-575-34	Sequence 34, Appl
44	6	4.0	211	1 US-08-383-619-16	Sequence 16, Appl
45	6	4.0	211	4 US-08-907-739-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-08-233-788A-59
Sequence 59, Application US/08233788A
Patent No. 5635617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Cloughier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
TITLE OF INVENTION: OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-Apr-1994
CLASSIFICATION: A35
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELE: 3723836 SEDANBERRY
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-233-788A-59
Query Match 80.8%, Score 122; DB 1; Length 151;
Best Local Similarity 100.0%; Pred. No. 1,1e-112;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 MKLLKVAFAIVVSGSALAGVPPQWGGGNNHGGSSGPDSTLSTIYOYGANALALQ 60
|||||

Db 1 MKLLKVAFAIIVSGSALAGVTPQMGCGGNHNGGSSGPDSTLSTIYOGSANAALALQ 60
QY 61 SPARKSETTITQSGYGADVGAGADNSTIELTQNGFRNNATIDQNAKNSDITVGYGG 120
Db 61 SPARKSETTITQSGYGADVGAGADNSTIELTQNGFRNNATIDQNAKNSDITVGYGG 120
QY 121 NN 122
Db 121 NN 122

RESULT 2
US-08-233-788A-57
; Sequence 57, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEEX: 3723836 SEDANBERRY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-233-788A-57

Query Match 74.2%; Score 112; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 5.9e-103;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 VVPQMGCGGNHNGGSSGPDSTLSTIYOGSANAALALQSDARKSETTITQSGYGADV 81
Db 1 VVPQMGCGGNHNGGSSGPDSTLSTIYOGSANAALALQSDARKSETTITQSGYGADV 60
QY 82 GCGADNSTIELTQNGFRNNATIDQNAKNSDITVGYGGNNALVNOTASDS 133
Db 61 GCGADNSTIELTQNGFRNNATIDQNAKNSDITVGYGGNNALVNOTASDS 112

RESULT 3
US-08-858-207A-291
; Sequence 291, Application US/08858207A
; Patent No. 5348328
; GENERAL INFORMATION:

APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Stodola, Robert
TITLE OF INVENTION: No. 6348328e1 Compounds
NUMBER OF SEQUENCES: 552
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,207A
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmli, Edward R.
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 291:
SEQUENCE CHARACTERISTICS:
LENGTH: 80 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6348328e
US-08-858-207A-291

Query Match 4.6%; Score 7; DB 4; Length 80;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 NAALALQ 60
Db 13 NAALALQ 19

RESULT 4
US-07-851-976B-8
; Sequence 8, Application US/07851976B
; Patent No. 5426043
; GENERAL INFORMATION:
; APPLICANT: De Graaff, Leendert H.
; APPLICANT: Visser, Jacob
; APPLICANT: Van Den Broeck, Henriette C.
; APPLICANT: Strozzyk, Francois
; APPLICANT: Kormelink, Felix J.M.
; APPLICANT: Boorman, Johannes C.P.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF ACETYL XILAN
; TITLE OF INVENTION: ESTERASES FROM FUNGAL ORIGIN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94104-2675

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/851,976B
FILING DATE: 19920316
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murashige, Kate H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 246152003200
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-813-5600
TELEFAX: 415-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-851-976B-8

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 304;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 32 HNGGNS 38
|||||
Db 115 HNGGNS 121

RESULT 5
US-08-291-609-8
Sequence 8, Application US/08291609
Patent No. 5681732
GENERAL INFORMATION:
APPLICANT: De Graaff, Leendert H.
APPLICANT: Visser, Jacob
APPLICANT: Van Den Broeck, Henriette C.
APPLICANT: Strozzyk, Francois
APPLICANT: Kormelink, Felix J.M.
APPLICANT: Boorman, Johannes C.P.
TITLE OF INVENTION: CLONING AND EXPRESSION OF ACETYL XYLAN
TITLE OF INVENTION: ESTERASES FROM FUNGAL ORIGIN
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94104-2675
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/291,609
FILING DATE: 17-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/851,976
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Murashige, Kate H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 246152003200
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-813-5600

TELEFAX: 415-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-291-609-8

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 304;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 32 HNGGNS 38
|||||
Db 115 HNGGNS 121

RESULT 6
US-08-401-136-8
Sequence 8, Application US/08401136
Patent No. 5763260
GENERAL INFORMATION:
APPLICANT: De Graaff, Leendert H.
APPLICANT: Visser, Jacob
APPLICANT: Van Den Broeck, Henriette C.
APPLICANT: Strozzyk, Francois
APPLICANT: Kormelink, Felix J.M.
APPLICANT: Boorman, Johannes C.P.
TITLE OF INVENTION: A METHOD TO ALTER THE PROPERTIES OF
TITLE OF INVENTION: ACETYLATED XYLAN
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 2000 PENNSYLVANIA AVENUE NW, SUITE 5500
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/401,136
FILING DATE: 08-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murashige, Kate H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 4615-0032.10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0764
TELEX: 90-4030
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-401-136-8

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 304;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 32 HNGGNS 38
|||||
Db 115 HNGGNS 121

RESULT 7
US-08-850-554-8
Sequence 8, Application US/08850554
Patent No. 6010892
GENERAL INFORMATION:
APPLICANT: De Graaff, Ieendert H.
APPLICANT: Visser, Jacob
APPLICANT: Van Den Broeck, Henriette C.
APPLICANT: Strozyk, Francois
APPLICANT: Kormelink, Felix J.M.
APPLICANT: Boorman, Johannes C.P.
TITLE OF INVENTION: A METHOD TO ALTER THE PROPERTIES OF
TITLE OF INVENTION: ACETYLATED XYLAN
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morison & Foerster
STREET: 2000 PENNSYLVANIA AVENUE NW, SUITE 5500
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/850,554
FILING DATE: 02-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/401,136
FILING DATE: 08-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Murashige, Kate H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 4615-0032.10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0764
TELEX: 90-4030
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-850-554-8

Query Match 4.6%; Score 7; DB 3; Length 304;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 HNGGNS 38
|||||||
DB 115 HNGGNS 121

RESULT 8
US-08-934-846-4
Sequence 4, Application US/08934846
Patent No. 5882898
GENERAL INFORMATION:
APPLICANT: Pearson, Stewart C.
APPLICANT: Greenwood, Rebecca C.
TITLE OF INVENTION: NOVEL FOIC
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA

COUNTRY: US
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/934,846
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dickinson, Todd O
REGISTRATION NUMBER: 28,354
REFERENCE/DOCKET NUMBER: GM10088
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2252
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 407 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-934-846-4

Query Match 4.6%; Score 7; DB 2; Length 407;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 NAALAQ 60
|||||||
DB 253 NAALAQ 259

RESULT 9
US-09-238-557-4
Sequence 4, Application US/09238557
Patent No. 6163472
GENERAL INFORMATION:
APPLICANT: Pearson, Stewart C.
APPLICANT: Greenwood, Rebecca C.
TITLE OF INVENTION: NOVEL FOIC
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/238,557
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/934,846
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dickinson, Todd O
REGISTRATION NUMBER: 28,354
REFERENCE/DOCKET NUMBER: GM10088
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2252

TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 407 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-238-557-4

Query Match 4.6%; Score 7; DB 4; Length 407;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 54 NAALALQ 60
|||||
DB 253 NAALALQ 259

RESULT 10

US-08-934-846-2
Sequence 2, Application US/08934846
Patent No. 5882898

GENERAL INFORMATION:
APPLICANT: Pearson, Stewart C.
APPLICANT: Greenwood, Rebecca C.
TITLE OF INVENTION: NOVEL FOIC
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:

ADDRESSEE: Dechert Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/934,846
FILING DATE:

CLASSIFICATION: 536
PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Dickinson, Todd Q

REGISTRATION NUMBER: 28,354
REFERENCE/DOCKET NUMBER: GM10088

TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2252

TELEFAX: 215-994-2222
TELEX:

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 418 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear

US-08-934-846-2

Query Match 4.6%; Score 7; DB 2; Length 418;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 54 NAALALQ 60
|||||
DB 253 NAALALQ 259

RESULT 11

US-09-238-557-2

Sequence 2, Application US/09238557
Patent No. 6165472

GENERAL INFORMATION:
APPLICANT: Pearson, Stewart C.
APPLICANT: Greenwood, Rebecca C.
TITLE OF INVENTION: NOVEL FOIC
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:

ADDRESSEE: Dechert Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/238,557
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/934,846
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Dickinson, Todd Q

REGISTRATION NUMBER: 28,354
REFERENCE/DOCKET NUMBER: GM10088

TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2252

TELEFAX: 215-994-2222
TELEX:

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 418 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear

US-09-238-557-2

Query Match 4.6%; Score 7; DB 4; Length 418;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 54 NAALALQ 60
|||||
DB 253 NAALALQ 259

RESULT 12

US-08-190-802A-97
Sequence 97, Application US/08190802A
Patent No. 5519003

GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
APPLICANT: Ron, Dorit
TITLE OF INVENTION: WD-40 - Derived Peptides and uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:

ADDRESSEE: Dehlinger & Associates
STREET: P.O. Box 60850
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-0850

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

RESULT 11

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/190,802A
FILING DATE: 01-FEB-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 8600-0139
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 97:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: CDC4 / CDC20 protein r1v, Fig. 15
US-08-190-802A-97

Query Match 4.0%; Score 6; DB 1; Length 29;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 IVVSGS 17
|11111|
DB 15 IVVSGS 20

RESULT 13
US-08-477-346-97
Sequence 97, Application US/08477346
Patent No. 6262023
GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 2000 Pennsylvania Avenue, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,346
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,072
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2550-0025.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
INFORMATION FOR SEQ ID NO: 97:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 amino acids
TYPE: amino acid

TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: CDC4 / CDC20 protein r1v, Fig. 15
US-08-477-346-97

Query Match 4.0%; Score 6; DB 4; Length 29;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 IVVSGS 17
|11111|
DB 15 IVVSGS 20

RESULT 14
US-08-473-089-97
Sequence 97, Application US/08473089
Patent No. 6342368
GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 2000 Pennsylvania Avenue, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,089
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2550-0025.22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
INFORMATION FOR SEQ ID NO: 97:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: CDC4 / CDC20 protein r1v, Fig. 15
US-08-473-089-97

Query Match 4.0%; Score 6; DB 4; Length 29;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 IVVSGS 17
|11111|
DB 15 IVVSGS 20

RESULT 15
US-09-053-197A-61

; Sequence 61, Application US/09053197A
; Patent No. 6022952
; GENERAL INFORMATION:
; APPLICANT: Weiner, Joel H.
; APPLICANT: Turner, Raymond J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PROTEIN
; TITLE OF INVENTION: SECRETION
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/053,197A
; FILING DATE: 01-APR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MacKnight, Kamrin T.
; REGISTRATION NUMBER: 38,230
; REFERENCE/DOCKET NUMBER: UALB-03293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-09-053-197A-61

Query Match 4.0%; Score 6; DB 3; Length 32;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 17 SALAGY 22
|11111
DB 19 SALAGY 24

Search completed: October 11, 2002, 21:21:11
Job time : 18 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 11, 2002, 21:17:28 ; Search time 20 Seconds
(without alignments)
725.474 Million cell updates/sec

Title: US-09-543-407-5

Perfect score: 151

Sequence: 1 MKLKVAFAAIVVSGSALA.....DSVWVRQVFGNNATANQY 151

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : PIR_71:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	151	100.0	151	2 JC6039	fimbria protein ag
2	151	100.0	151	2 AJ0635	major curlin chain
3	151	100.0	151	2 S70788	curlin protein cs9
4	11	7.3	152	2 D90806	curlin major subun
5	11	7.3	152	2 H85665	hypothetical prote
6	8	5.3	436	2 T36490	probable transmem
7	8	5.3	439	2 AF1813	hypothetical prote
8	5.3	552	2 T34611	2	NADH dehydrogenase
9	7	4.6	62	2 T36447	hypothetical prote
10	7	4.6	145	2 D89920	conserved hypothet
11	7	4.6	238	2 I48605	insulin-like growt
12	7	4.6	251	2 AH2344	ATP-binding protei
13	7	4.6	254	2 AE0940	probable ribulose-
14	7	4.6	276	2 AI3572	oligopeptide trans
15	7	4.6	284	2 A84682	probable zinc-fing
16	7	4.6	314	2 D64762	2,3-dihydroxypheny
17	7	4.6	314	2 C96799	2,3-dihydroxypheny
18	7	4.6	314	2 G85529	2,3-dihydroxypheny
19	7	4.6	317	2 C85432	hypothetical prote
20	7	4.6	323	2 F86192	hypothetical prote
21	7	4.6	325	2 T23374	hypothetical prote
22	7	4.6	347	2 H65504	hypothetical prote
23	7	4.6	362	2 AF1950	hypothetical prote
24	7	4.6	362	2 AD1315	anthranilate phosph
25	7	4.6	367	2 AD1687	prephenate dehydro
26	7	4.6	387	2 B70610	hypothetical prote
27	7	4.6	388	2 T15075	hypothetical prote
28	7	4.6	395	2 T10828	translation elonga
29	7	4.6	408	2 S62725	translation elonga

30	7	4.6	413	2 S61305	cych protein - Par
31	7	4.6	416	2 H95022	dihydrofolate synt
32	7	4.6	418	2 B97894	dihydrofolate synt
33	7	4.6	429	2 B83257	folypolyglutamate
34	7	4.6	436	2 T36432	probable transmem
35	7	4.6	458	2 E82340	probable gluconate
36	7	4.6	545	2 A46281	tetrahydrofolylpol
37	7	4.6	585	2 S46825	hypothetical prote
38	7	4.6	587	2 I49357	tetrahydrofolylpol
39	7	4.6	587	2 S65755	tetrahydrofolylpol
40	7	4.6	614	2 D84191	dolichol-P-glucose
41	7	4.6	652	2 E97857	cell surface anti
42	7	4.6	796	2 T08555	pathogenesis-relat
43	7	4.6	951	2 T08987	probable cadmium-t
44	7	4.6	982	2 T19526	hypothetical prote
45	6	4.0	35	2 B33770	hypothetical prote

ALIGNMENTS

RESULT 1

JC6039 fimbria protein agfa precursor - Salmonella enteritidis
C:Species: Salmonella enteritidis
C:Date: 31-Dec-1996 #sequence-revision 31-Dec-1996 #text-change 08-Oct-1999
C:Accession: JC6039; PC6015; A44898
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.

J. Bacteriol. 178, 662-667, 1996
A:Title: Salmonella enteritidis agfaBC operon encoding thin, aggregative fimbriae.
A:Reference number: JC6039; MUID:96146512

A:Accession: JC6039

A:Molecule type: DNA

A:Residues: 1-151 <COI>

A:Cross-references: GB:043280; NID:91184712; PIDN:AAC43599.1; PID:91184714

A:Accession: PC6015

A:Molecule type: Protein

A:Residues: 21-52 <CO2>

A:Experimental source: strain 27655-3b

R:Collinson, S.K.; Emdy, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.

J. Bacteriol. 173, 4773-4781, 1991

A:Title: Purification and characterization of thin, aggregative fimbriae from Salmone

A:Reference number: A44898; MUID:91310586

A:Accession: 27655

A:Status: Preliminary

A:Molecule type: Protein

A:Residues: 21-33 <CO3>

A>Note: sequence extracted from NCBI backbone (NCBI:45936)

C:Genetics:

A:Gene: agfa

C:Function:

A:Description: major component of thin aggregative fimbriae

A>Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator

C:Keywords: fimbria

F.1-20/Domain: signal sequence #status predicted <SIG>

F.21-151/Product: fimbria protein agfa #status experimental <MAT>

Query Match Best Local Similarity 100.0%; Score 151; DB 2; Length 151; Pred. No. 1.3e-146; Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLKVAFAAIVVSGSALAGVPPQGGGNGGNSGPPSTLSTIYOGSANAALAQ 60
DB 1 MKLKVAFAAIVVSGSALAGVPPQGGGNGGNSGPPSTLSTIYOGSANAALAQ 60
QY 61 SPARKSETTITSGYNGADVGAGADNSTLETONGFRNNATIDQWNAKNSDITVGOYGC 120
DB 61 SPARKSETTITSGYNGADVGAGADNSTLETONGFRNNATIDQWNAKNSDITVGOYGC 120
QY 121 NNAAIVNOTASDSVWVRQVFGNNATANQY 151
DB 121 NNAAIVNOTASDSVWVRQVFGNNATANQY 151

Db 121 NNAALVNQTASDSSVMVRQVGFGNNATANQY 151

RESULT 2

A10635
major_curlin_chain_precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A>Note: This species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C:Accession: A10635
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.S.; Connerton, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, M.; Hingle, B.; Jones, S.; Kaul, R.; MacArthur, C.T.; Maitland, N.; Miller, J.; Parkhill, J.; Quail, M.A.; Sehra, V.; Skellern, J.; Smith, R.; Stalker, J.; Young, L.; Brown, K.E.; de Lencastre, E.; et al.: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium DT104. Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; et al.: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium DT104. Nature 413, 848-852, 2001
A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium DT104.
A:Reference number: AB0502; PMID:11677608
A:Accession: A10635
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-151 <PAR>
A:Cross-references: GB:A1513382; PIDN:CAD08268.1; PID:g16502315; GSPDB:GN00176
C:Genetics:
A:Gene: STY1181

Query Match	100.0%;	Score 151;	DB 2;	Length 151;
Best Local Similarity	100.0%;	Pred. No. 1.3e-146;		
Matches 151;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0

QY 1 MKLLKVAFAAIVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALDQ 60

Db 1 MKLLKVAFAIVSSGALAGVVPQWGGGNNHNGGGNSGPDLSIIQIGSANNALRDY

[illegible][illegible]

07 121 ANNAHVSSTSSSYMPVOEGENATANY 15

RESULT 3

5/0/00
curlin protein csga precursor - Escherichia coli
N; Alternate names: csga protein; major curlin protein

C:Species: *Escherichia coli*
C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 21-Jul-2000
C:Accession: S70788 G64846; S31202; S34560; S34559
R:Hemmar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A:Title: Expression of two csq operons is required for production of fibronectin- and C
A:Reference number: S70783; MUID:56414468
A:Accession: S70788
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <HAM>
A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62282.1; PID:g1147564
A:Experimental source: strain K12, substrain W3110
A:Note: the nucleotide sequence was submitted to the EMBL Data library, August 1995
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.: Rose, D.J.; Mau, B.; Snao, Y.
Science 277, 1453-1462, 1997

A:Title: The RPOS sigma factor relieves H-NS-mediated transcriptional repression of *hns*
A:Reference number: S31202; MUID:93211294

Query Match	7.3%;	Score 11;	DB 2;	Length 151;
Best Local Similarity	100.0%;	Pred. No. 0.0017;		
Matches 11;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY 15 SG S A L A G V P Q 25

Db . 15 SCSALAGVPQ 25

RESULT 4

D90806 curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain R)

C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: D90806
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
gasaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: D90806
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-152 <MAX>
A;Cross-references: GB:BA000007; PIDN:BA34843.1, PID:q13360880; GSPDB:GN00154
C;Experimental source: strain O157:H7, substrain RMD 050952
C;Genetics:
A;Gene: ECs1420

Query Match
Post Local Similarity

Matches 11; Conserve

27

Db 15 SSSALAGVVPQ 25

A:Reference number: A64720; MUID:97428617
A:Accession: G64846
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <BAT>
A:Cross-references: GB:A000205; GI:000096; NID:g1787265; PIDN:AAC74126.1; PID:g1787279;
A:Experimental source: strain K-12, substrain MG1655
R.Olsen, A.; Arngvist, A.; Hammar, M.; Sukhopolt, S.; Normark, S.
Mol. Microbiol. 7, 523-536, 1993

RESULT 5
H85665
hypothetical protein csgA [imported] - Escherichia coli (strain O15:H7, substrain EDL)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: H85665
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Ma
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimlant, E.; Potamousts, K.; Apod

Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; PMID:21074935; PMID:11206551
A:Accession: H85655
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <STO>
A:Cross-references: GB:AE005174; NID:912514574; PIDN:ANG55788.1; GSPDB:GN00145; UMG:Z16
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: cs9a

Query Match 7.3%; Score 11; DB 2; Length 152;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 SGSLAGVVP 25
|||||
Db 15 SGSLAGVVP 25

RESULT 6

T36490
probable transmembrane transport protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: T36490
R:Saunders, D.C.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21608
A:Accession: T36450
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-436 <SAU>
A:Cross-references: EMBL:AL096822; PIDN:CAB46932.1; GSPDB:GN00070; SCOEDB:SCG03.10C
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCG03.10C
C:Superfamily: bicyclomycin resistance protein

Query Match 5.3%; Score 8; DB 2; Length 436;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 GSALAGV 23
|||||
Db 182 GSALAGV 189

RESULT 7

AF1813
hypothetical protein alr0054 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AF1813
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matsuoka, A.; Iriyuchi,
N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; PMID:21595285; PMID:11759840
A:Accession: AF1813
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-439 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA877578.1; PID:917135032; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr0054

Query Match 5.3%; Score 8; DB 2; Length 439;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 SALAGVVP 24
|||||
Db 363 SALAGVVP 370

RESULT 8

T34611
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 (similarity) - Streptomyces coel
C:Species: Streptomyces coelicolor
C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-May-2000
C:Accession: T34611
R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream,
submitted to the EMBL Data Library, June 1999
A:Reference number: Z21549
A:Accession: T34611
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-552 <SAU>
A:Cross-references: EMBL:AL078618; PIDN:CAB44518.1; GSPDB:GN00070; SCOEDB:SC10A7.08C
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: nuon; SCOEDB:SC10A7.08C
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 2
C:Keywords: membrane-associated complex; NAD; oxidoreductase

Query Match 5.3%; Score 8; DB 2; Length 552;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 VAAFAAIV 13
|||||
Db 71 VAAFAAIV 78

RESULT 9

T36447
hypothetical protein SCF43A.24c - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36447
R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21598
A:Accession: T36447
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-62 <SEB>
A:Cross-references: EMBL:AL096837; PIDN:CAB48911.1; GSPDB:GN00070; SCOEDB:SCF43A.24C
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCF43A.24C

Query Match 4.6%; Score 7; DB 2; Length 62;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 55 AALAIQS 61
|||||
Db 39 AALAIQS 45

RESULT 10

D89920
conserved hypothetical protein SA1261 [imported] - Staphylococcus aureus (strain N315
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: D89920
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-Oi, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsuka, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146
 A:Accession: D89920
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-145 <KUR>
 A:Cross-references: GB:BA000018; PID:g13701226; PIDN:BA842521.1; GSPDB:GN00149
 A:Experimental source: strain N315
 C:Genetics:
 A:Gene: SA1261

Query Match 4.6%; Score 7; DB 2; Length 145;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 91 ELRONGF 97
 |||||
 Db 18 ELRONGF 24

RESULT 11

insulin-like growth factor binding protein-6 - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 31-Jan-2000
 C:Accession: I48605
 R:Schuller, A.G.; Groffen, C.; Van Neck, J.W.; Zwartthoff, E.C.; Drop, S.L.
 Mol. Cell. Endocrinol. 104, 57-66, 1994
 A:Title: cDNA cloning and mRNA expression of the six mouse insulin-like growth factor b
 A:Reference number: I48600; MUID:95121750
 A:Accession: I48605
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-238 <RES>
 A:Cross-references: EMBL:X81584; NID:g550386; PIDN:CA557274.1; PID:g550387
 C:Genetics:
 A:Gene: IGFBP-6
 C:Superfamily: insulin-like growth factor binding protein 1; thyroglobulin type I repeat
 F:160-232/Domain: thyroglobulin type I repeat homology <TRY1>

Query Match 4.6%; Score 7; DB 2; Length 238;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 SSGSALAG 21
 |||||
 Db 23 SSGSALAG 29

RESULT 12

AH2344
 ATP-binding protein of ABC transporter alr4311 [imported] - Anabaena sp. (strain PCC 712
 C:Species: Anabaena sp.
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
 C:Accession: AH2344
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi
 Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete genomic sequence of the filamentous Nitrogen-fixing Cyanobacterium Ana
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AH2344
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-251 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA876010.1; PID:g17133447; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: alr4311

Query Match 4.6%; Score 7; DB 2; Length 251;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 125 LVNOTAS 131
 |||||
 Db 222 LVNOTAS 228

RESULT 13

AE0940
 probable ribulose-phosphate 3-epimerase STY3790 [imported] - Salmonella enterica subs
 C:Species: Salmonella enterica subsp. enterica serovar Typhi
 A:Note: this species has also been called Salmonella typhi
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
 C:Accession: AE0940
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
 th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.W.; Dowd, L.; White, N.; Farr
 S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
 A:Reference number: AB0502; PMID:11677608
 A:Accession: AE0940
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-254 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD09543.1; PID:g16504659; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY3790

Query Match 4.6%; Score 7; DB 2; Length 254;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 VVGSAL 19
 |||||
 Db 218 VVGSAL 224

RESULT 14

AI3572
 oligopeptide transport system permease protein oppC [imported] - Brucella melitensis
 C:Species: Brucella melitensis
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
 C:Accession: AI3572
 R:DelVecchio, V.G.; Kaputral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanov
 Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit
 A:Reference number: AD3252; PMID:11756688
 A:Accession: AI3572
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-276 <KUR>
 A:Cross-references: GB:AE08918; PIDN:AL53748.1; PID:g17984674; GSPDB:GN00191
 A:Experimental source: strain 16M
 C:Genetics:
 A:Gene: BMRI10506
 A:Map position: 11
 C:Superfamily: oligopeptide permease protein oppB

Query Match 4.6%; Score 7; DB 2; Length 276;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 15

AB4682
 probable zinc-finger protein [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
 C:Accession: AB4682

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Talion, L.;
euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
A:Reference number: A04420; MUID:20083487
A:Accession: A04682
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-284 <STO>
A:Cross-references: GB:AE002093; NID:94803961; PIDN:AAD29833.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g28200
A:Map position: 2
C:Superfamily: Arabidopsis thaliana hypothetical protein F12E4.290

Query Match 4.6%; Score 7; DB 2; Length 284;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 VSGSALA 20
|||||||
DB 168 VSGSALA 174

Search completed: October 11, 2002, 21:20:42
Job time : 22 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 11, 2002, 21:14:58 ; Search time 11 Seconds
(without alignments)
531.514 Million cell updates/sec

Title: US-09-543-407-5
Perfect score: 151
Sequence: 1 MKLKVAFALIVSGSALA.....DSSVMRVGVGNNATANQY 151

Scoring table: OLIGO
Gapop 60.0 , Gapept 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Swissprot_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	151	100.0	151	1	CSGA_SALTY
2	11	7.3	151	1	P28307 escherichia
3	7	4.6	118	1	RL20_PSEFL
4	7	4.6	118	1	RL20_PSEFL
5	7	4.6	238	1	IBP6_MOUSE
6	7	4.6	280	1	VT24_AGRYT
7	7	4.6	314	1	MHPB_ECOLI
8	7	4.6	395	1	EFTU_CYTLY
9	7	4.6	408	1	EFTU_CHACO
10	7	4.6	585	1	YH80_YEAST
11	7	4.6	587	1	FOLC_HUMAN
12	7	4.6	600	1	FOLC_MOUSE
13	7	4.6	600	1	RM12_MOUSE
14	7	4.6	796	1	PRH_ARATH
15	7	4.6	833	1	HMDH_AGRIP
16	7	4.6	951	1	AHM3_ARATH
17	6	4.0	86	1	PTHP_STRMU
18	6	4.0	87	1	PTHP_STRBO
19	6	4.0	87	1	PTHP_STRSL
20	6	4.0	87	1	VAF2_BACLI
21	6	4.0	88	1	PTHP_LACCA
22	6	4.0	88	1	PTHP_LACCA
23	6	4.0	88	1	PTHP_LACCA
24	6	4.0	88	1	PTHP_LACCA
25	6	4.0	88	1	PTHP_LACCA
26	6	4.0	88	1	PTHP_LACCA
27	6	4.0	146	1	AP4A_MOUSE
28	6	4.0	148	1	FLAV_DESDE
29	6	4.0	153	1	Y230_METTA
30	6	4.0	158	1	YU95_MYCTU
31	6	4.0	166	1	CHED_BACSU
32	6	4.0	174	1	ATPF_ENTHR
33	6	4.0	176	1	ATPF_GALSU

34	6	4.0	203	1	CHPE_PSEAE
35	6	4.0	210	1	UPP_DEIRA
36	6	4.0	211	1	LEXA_XYTRA
37	6	4.0	226	1	BIOD_XYTRA
38	6	4.0	226	1	GPH_VIBCH
39	6	4.0	235	1	RS4E_THEAC
40	6	4.0	235	1	YD22_ECOLI
41	6	4.0	236	1	UNG_CHLPN
42	6	4.0	240	1	PYRH_AQUAE
43	6	4.0	244	1	CBR2_MOUSE
44	6	4.0	247	1	SNAB_MOUSE
45	6	4.0	248	1	PCN2_SULOH

ALIGNMENTS

RESULT 1	ID	CSGA_SALTY	STANDARD:	PRT:	151 AA.
AC	P55225:				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DE	Major curlin subunit precursor (Fimbrin SEF17).				
GN	CSGA OR AGPA OR STM1144 OR STY1181.				
OS	Salmonella typhimurium,				
OS	Salmonella typhimurium,				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC	Salmonella.				
OX	NCBI_Taxid=602, 601, 592;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S. typhimurium; STRAIN=SR-11;				
RX	MEDLINE=98117058; PubMed=9457880;				
RA	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;				
RT	"Curli fibers are highly conserved between Salmonella typhimurium and				
RT	Escherichia coli with respect to operon structure and regulation.";				
RL	J. Bacteriol. 180:722-731(1998).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S. typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;				
RX	MEDLINE=21534948; PubMed=11677609;				
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,				
RA	Leontiney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,				
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Muiyaney E.,				
RA	Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,				
RA	Waterston R., Wilson R.K.;				
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium				
RT	LT2.";				
RL	Nature 413:852-856(2001).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S. typhi; STRAIN=CT18;				
RX	MEDLINE=21534947; PubMed=11677608;				
RA	Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,				
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebalina M.,				
RA	Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,				
RA	Crofton A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,				
RA	Fellwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,				
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,				
RA	Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,				
RA	Whitehead S., Barrett B.G.;				
RT	"Complete genome sequence of a multiple drug resistant Salmonella				
RT	enterica serovar Typhi CT18.";				
RL	Nature 413:848-852(2001).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S. enteritidis; STRAIN=27655-3B;				
RX	MEDLINE=96146512; PubMed=8550497;				
RA	Collinson S.K., Clouthier S.C., Doran J.L., Bansen P.A., Kay W.W.;				
RT	"Salmonella enteritidis agfBAC operon encoding thin, aggregative				

RT	fimbriae.";
RL	Bacterial. 178:662-667(1996).
RN	151
RP	SEQUENCE OF 21-151 FROM N.A.
RC	SPECIES-S.entertitidis; STRAIN=27655-3B;
RX	MEDLINE=N4013373; PubMed=8104955;
RA	Doran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,
RB	Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;
RT	"DNA-based diagnostic tests for Salmonella species targeting agfa,
RU	the structural gene for thin, aggregative fimbriae."
RV	J. Clin. Microbiol. 31:2263-2273(1993).
RN	[6]
RP	SEQUENCE OF 21-33.
RC	SPECIES-S.entertitidis; STRAIN=27655-3B;
RX	MEDLINE=91310586; PubMed=1677357;
RA	Collinson S.K., Emsedy L., Mueller K.-M., Trust T.J., Kay W.W.;
RB	Purification and characterization of thin, aggregative fimbriae from
RT	Salmonella enteritidis."
RU	J. Bacteriol. 173:4773-4781(1991).
CC	-I- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC	COLLED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC	TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC	FIBRONECTIN.
CC	-I- SIMILARITY: STRONG, TO E.COLI CSGA.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC	or send an email to license@isb-sdb.ch).
CC	-----
DR	EMBL; AJ002301; CA05317.1; -
DR	EMBL; AE008749; AAL20074.1; -
DR	EMBL; AL627263; CAD08268.1; -
DR	EMBL; U43280; AAC43599.1; -
ST	StyGene; SG10608; csqa.
FT	Fimbrin; Signal; Complete proteome.
KW	SIGNAL
FT	CHAIN 1 20
FT	CONFLICT 21 151 MAJOR CURLIN SUBUNIT.
FT	CONFICT 134 151 SVYVQVGFGNNATANQY -> DSYTOVAS (IN
FT	REF. 5)
SQ	SEQUENCE 151 AA: 15305 MW: B7DACD16B621359 CRC64:
Query Match	100.0%; Score 151; DB 1; Length 151;
Best Local Similarity	100.0%; Pred. No. 6.8e-144;
Matches 151; Conservative	0; Mismatches 0; Indels 0; Gaps 0.
OY	1 MKLLKAAAFRAIVSSSALAGVPYPWGSGGNHNGGNSGPDSLSITGYGSANAALALO 60
DB	1 MKLLKAAFAIAIVSSSALAGVYPWGSGGNHNGGNSGPDSLSITGYGSANAALALO 60
OY	61 SDARKSEPTTTGGCGYGADVGAGANSTIELQNQFERNNAITIDQNNAKNSDITTYGOYGG 120
DB	61 SDARKSETTTTGGCGYGADVGAGADNSTIELQNQFRNNAITIDQNNAKNSDITTYGOYGG 120
OY	121 NNAALVNOTASDSSVMRYOYGFEGNNATANQY 151
DB	121 NNAALVNOTASDSSVMRYOYGFEGNNATANQY 151
RESULT 2	
CSGA_ECOLI	STANDARD; PRT; 151 AA.
ID	CSGA_ECOLI
AC	P28307;
DT	01-DEC-1992 (Rel. 24, Created)
DT	01-OCT-1996 (Rel. 34, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Major curlin subunit precursor.
GN	CSGA OR BI042.
OS	Escherichia coli.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

CC Escherichia.
 CC NCBI_TaxID=562;
 CC [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=K12 / W3110;
 CC RX MEDLINE=93211294; PubMed=8459772;
 CC RA Olsen A., Arngvist A.;
 CC RT "The Rpos sigma factor relieves H-NS-mediated transcriptional
 CC repression of csgA, the subunit gene of fimbriectin-binding curli in
 CC Escherichia coli.";
 CC RL Mol. Microbiol. 7:523-536(1993).
 CC [2]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=K12 / MC4100;
 CC RX MEDLINE=96414468; PubMed=8817489;
 CC RA Hammar W., Arngvist A., Bian Z., Normark S.;
 CC RT "Expression of two csg operons is required for production of
 CC fimbriectin- and Congo red-binding curli polymers in Escherichia coli
 CC K-12.";
 CC RL Mol. Microbiol. 18:661-670(1995).
 CC [3]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=K12 / MG1655;
 CC RX MEDLINE=97426617; PubMed=9276503;
 CC RA Blattner F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
 CC RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,
 CC RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 CC RA Mau B., Shao Y.;
 CC RT "The complete genome sequence of Escherichia coli K-12.";
 CC RL Science 277:1453-1474(1997).
 CC [4]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=K12;
 CC RX MEDLINE=97061202; PubMed=8905232;
 CC RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 CC RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 CC RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 CC RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 CC RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 CC RA Yano M., Horiiuchi T.;
 CC RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 CC corresponding to the 12.7-28.0 min region on the linkage map.";
 CC RL DNA Res. 3:137-155(1996).
 CC [5]
 CC RP SEQUENCE OF 21-40.
 CC RC STRAIN=K12 / YMEJ;
 CC RX MEDLINE=93023873; PubMed=1357528;
 CC RA Arngvist A., Olsen A., Pfeiffer J., Russell D.G., Normark S.;
 CC RT "The Crl protein activates cryptic genes for curli formation and
 CC fimbriectin binding in Escherichia coli HB101.";
 CC RL Mol. Microbiol. 6:2443-2452(1992).
 CC [6]
 CC RP SEQUENCE OF 21-31.
 CC RX MEDLINE=91310586; PubMed=1677357;
 CC RA Collinson S.K., Emoeely L., Trust T.J., Kay W.W.;
 CC RT "Purification and characterization of thin, aggregative fimbriae from
 CC Salmonella enteritidis.";
 CC RL J. Bacteriol. 173:4773-4781(1991).
 CC [7]
 CC RP FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN.
 CC [8]
 CC RP SIMILARITY: STRONG, TO SALMONELLA CGSA.
 CC [9]
 CC RP SIMILARITY: TO CURLIN MINOR SUBUNIT (CSGB).
 CC [10]
 CC RP This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation on
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC [11]

DR EMBL; L04979; AAA23616.1; -
 DR EMBL; X90754; CAA62282.1; -
 DR EMBL; AE000205; AAC74126.1; -
 DR EMBL; D90741; BAA35832.1; -
 DR EMBL; D90742; BAA35840.1; -
 DR Ecogene; EG11489; cs9a.
 KW Fimbrtia; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 151
 FT CONFLICT 7 7 A -> E (IN REF. 1).
 SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;

Query Match 7.3%; Score 11; DB 1; Length 151;
 Best Local Similarity 100.0%; Pred. No. 0.00095;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 SGSLAGVVPQ 25
 DB 15 SGSLAGVVPQ 25

RESULT 3

RL20_PSEFL STANDARD; PRT; 118 AA.
 AC Q9X6E8;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 50S ribosomal protein L20.
 GN RPLT.
 OS Pseudomonas fluorescens.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 CC Pseudomonas.
 CC NCBI_TaxID=294;
 OX RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=CHA0;
 RA Blumer C., Heeb S., Pessi G., Haas D.;
 RT "Global Gaca-steered control of secondary metabolism in Pseudomonas
 RT fluorescens acts on specific ribosome binding sites."
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND IS
 CC NECESSARY TO THE IN VITRO ASSEMBLY PROCESS OF THE 50S RIBOSOMAL
 CC SUBUNIT: IT IS NOT INVOLVED IN THE PROTEIN SYNTHESIZING FUNCTIONS
 CC OF THAT SUBUNIT (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE L20P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF136400; AAD34789.1; -
 DR InterPro; IPR001081; Ribosomal_L20.
 DR Pfam; PF00453; Ribosomal_L20; 1.
 DR PRINTS; PR00062; RIBOSOMAL_L20.
 DR PRODOM; PD002389; Ribosomal_L20; 1.
 DR PROSITE; PS00937; RIBOSOMAL_L20; 1.
 KW Ribosomal protein; rRNA-binding.
 SO SEQUENCE 118 AA; 13295 MW; E83E575681EE452 CRC64;

Query Match 4.6%; Score 7; DB 1; Length 118;
 Best Local Similarity 100.0%; Pred. No. 7.8;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 AAFPAIV 13
 DB 104 AAFPAIV 110

RESULT 4
 ID RL20_PSEFL STANDARD; PRT; 118 AA.
 AC P52828;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 50S ribosomal protein L20.
 GN RPLT.
 OS Pseudomonas syringae (pv. syringae)
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 CC Pseudomonas.
 CC NCBI_TaxID=321;
 OX RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=SUPP27;
 RA MEDLINE=96196155; PubMed=8626280;
 RA Kitten T., Willis D.K.;
 RT "Suppression of a sensor kinase-dependent phenotype in Pseudomonas
 RT syringae by ribosomal proteins L35 and L20."
 RL J. Bacteriol. 178:1548-1555(1996).

CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND IS
 CC NECESSARY TO THE IN VITRO ASSEMBLY PROCESS OF THE 50S RIBOSOMAL
 CC SUBUNIT: IT IS NOT INVOLVED IN THE PROTEIN SYNTHESIZING FUNCTIONS
 CC OF THAT SUBUNIT (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE L20P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U44118; AAB05016.1; -
 DR InterPro; IPR001081; Ribosomal_L20.
 DR Pfam; PF00453; Ribosomal_L20; 1.
 DR PRINTS; PR00062; RIBOSOMAL_L20.
 DR PRODOM; PD002389; Ribosomal_L20; 1.
 DR PROSITE; PS00937; RIBOSOMAL_L20; 1.
 KW Ribosomal protein; rRNA-binding.
 SO SEQUENCE 118 AA; 13314 MW; 2BF171F5681EE443 CRC64;

Query Match 4.6%; Score 7; DB 1; Length 118;
 Best Local Similarity 100.0%; Pred. No. 7.8;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 AAFPAIV 13
 DB 104 AAFPAIV 110

RESULT 5

IBP6_MOUSE STANDARD; PRT; 238 AA.
 AC P47860;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Insulin-like growth factor binding protein 6 precursor (IGFBP-6)
 DE (IBP-6) (IGF-binding protein 6).
 GN IGFBP6 OR IGFBP-6.
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=Liver;
 RX MEDLINE=95121750; PubMed=7529732;
 RA Schuller A.G.P., Groffen C., van Neck J.W., Zwarthoff E.C.,
 RA Drop S.L.S.;

```

RT "cDNA cloning and mRNA expression of the six mouse insulin-like
RT growth factor binding proteins."
RL Mol. Cell. Endocrinol. 104:57-66(1994).
CC -1- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFs
CC AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
CC PROMOTING EFFECTS OF THE IGFs ON CELL CULTURE. THEY ALTER THE
CC INTERACTION OF IGFs WITH THEIR CELL SURFACE RECEPTORS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- PFM: O-GLYCOSYLATED.
CC -1- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X81584; CA57274.1; -.
DR HSSP; P24593; IBOE.
DR MGD; MGI:96441; Igfbp6.
DR InterPro: IPR000867; IGFBP.
DR InterPro: IPR000716; thyroglobulin_1.
DR Pfam; PF00219; IGFBP_1.
DR Pfam; PF00086; thyroglobulin_1; 1.
DR SMART; SM00211; IB; 1.
DR SMART; SM00211; TY; 1.
DR PROSITE; PS00222; IGF_BINDING; FALSE_NEG.
DR PROSITE; PS00484; THYROGLOBULIN_1; 1.
DR Growth factor binding; Signal; Glycoprotein.
KM SIGNAL 1 25
FT CHAIN 26 238
FT DOMAIN 184 232
FT DISULFID 30 33
FT DISULFID 41 45
FT DISULFID 58 64
FT DISULFID 72 85
FT DISULFID 79 105
FT DISULFID 160 188
FT DISULFID 199 210
FT DISULFID 212 232
SQ SEQUENCE 238 AA; 25402 MW; CCECDICD6AD9559F9 CRC64;

Query Match 4.6%; Score 7; DB 1; Length 238;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 SSSALAG 21
DB 23 SSSALAG 29

```

```

RA Salomone J.-Y., Crouzet P., de Ruffray P., Otten L.;
RT "Characterization and distribution of tarrate utilization genes in
RT the grapevine pathogen Agrobacterium vitis."
RL Mol. Plant Microbe Interact. 9:401-408(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32375; AAB61632.1; -.
DR Hypothetical protein; Plasmid.
SQ SEQUENCE 280 AA; 31571 MW; 89050A6A81A5A61 CRC64;

Query Match 4.6%; Score 7; DB 1; Length 280;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 SSSALAG 21
DB 182 SSSALAG 188

```

```

RESULT 6
YT24_AGRVI STANDARD; PRT; 280 AA.
ID YT24_AGRVI
AC P70796;
DR 15-DEC-1998 (Rel. 37, Created)
DR 15-DEC-1998 (Rel. 37, Last sequence update)
DR 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 31.6 kDa protein in TAR-1 TTUC' 3' region (ORF24).
OS Agrobacterium vitis (Rhizobium vitis).
OG Plasmid pTIRAB3.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
CC Rhizobiaceae; Rhizobium.
CX NCBI_TaxID=373;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=AB3;
RX MEDLINE=96252899; PubMed=8672817;

```

```

RESULT 7
MHPB_ECOLI STANDARD; PRT; 314 AA.
ID MHPB_ECOLI
AC P54711; P77461; P77048;
DR 01-OCT-1996 (Rel. 34, Created)
DR 01-NOV-1997 (Rel. 35, Last sequence update)
DR 01-MAR-2002 (Rel. 41, Last annotation update)
DE 2,3-dihydroxyphenylpropionate 1,2-dioxygenase (EC 1.13.11.-).
GN MHPB OR B0348 OR Z0446 OR EC50403.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
CX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=96359381; PubMed=8752345;
RA Spence E.L., Kawamukai M., Sanvoisin J., Braven H., Bugg T.D.H.;
RT "Catechol dioxygenases from Escherichia coli (MhpB) and Alcaligenes
RT eutrophus (MpcI): sequence analysis and biochemical properties of a
RT third family of extradiol dioxygenases."
RT J. Bacteriol. 178:5249-5256(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / CS520;
RA Ferrandez A., Garcia J.L., Diaz E.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RA Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdt O., Lew H., Lin D.,
RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-0157:H7 / ED933 / ATCC 700927;
RA MEDLINE-21074935; PubMed-11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boulton A., Shao Y., Miller L.,
RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoculis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN-0157:H7 / RIMD 0509952;
RA MEDLINE-21156231; PubMed-11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli"
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [7]
RP SEQUENCE OF 1-15. AND CHARACTERIZATION.
RX MEDLINE-94002258; PubMed-8399388;
RA Bugy T.D.H.;
RT "Overproduction, purification and properties of 2,3-
RT dihydroxyphenylpropionate 1,2-dioxygenase from Escherichia coli.";
RL Biochim. Biophys. Acta 1202:258-264(1993).
CC -1- FUNCTION: EXTRADIOL CLEAVAGE OF 2,3-DIHYDROXYPHENYLPROPIONIC ACID.
CC -1- COFACTOR: FERROUS ION.
CC -1- PATHWAY: 3-HYDROXYPHENYLPROPIONATE DEGRADATION PATHWAY.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SIMILARITY: TO A. EUROPHUS MCPI.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; D86329; BAA13053.1; -.
CC DR EMBL; Y09555; CA70748.1; -.
CC DR EMBL; AE000142; AAC73451.1; -.
CC DR EMBL; U73857; AAB18072.1; -.
CC DR EMBL; AE005214; AAC54699.1; -.
CC DR EMBL; AF002551; BAB33826.1; -.
CC DR EcoGene; EG20274; mhpB.
CC DR InterPro: IPR004183; L19B.
CC DR Pfam; PF02900; L19B.1.
CC KW Oxidoreductase; Dioxygenase; Aromatic hydrocarbons catabolism; Iron:
CC Repeat; Complete proteome.
FT DOMAIN 1 171 1.
FT DOMAIN 2 172 314 2.
FT DOMAIN 3 138 140 2.
FT CONFLICT 152 157 A -> H (IN REF. 1).
FT CONFLICT 157 157 R -> T (IN REF. 1).
SQ SEQUENCE 314 AA; 34196 MW; EID5AB574E5DFE05 CRC64;

Query Match 4.6%; Score 7; DB 1; Length 314;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 VAAFAAI 12
DB 276 VAAFAAI 282

RESULT 8
EFTU_CYTLY STANDARD: PRT; 395 AA.
AC P42474;

```

```

DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Elongation factor Tu (EF-Tu).
GN TUF.
OS Cytophaga lytica.
OC Bacteria; CFb group; Flavobacteriia; Flavobacteriaceae; Cellulophaga.
OC NCBI_TaxID=979;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 2039;
RA MEDLINE-94368062; PubMed-8085791;
RA Ludwig W., Neumaler J., Klugbauer N., Brockmann E., Roller C.,
RA Klugbauer S., Reetz K., Schachtner I., Ludvigsen A.,
RA Bachleitner M., Fischer U., Schleifer K.H.;
RT "Phylogenetic relationships of Bacteria based on comparative sequence
RT analysis of elongation factor Tu and Arp-synthase beta-subunit
RT genes.";
RL Antoine Van Leeuwenhoek 64:285-305(1993).
CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X77035; CA54324.1; -.
CC DR HSSP; P02990; IEFU.
CC DR InterPro: IPR000795; GTP_EFTU.
CC DR InterPro: IPR004161; GTP_EFTU_D2.
CC DR InterPro: IPR004160; GTP_EFTU_D3.
CC DR Pfam; PF00009; GTP_EFTU; 1.
CC DR Pfam; PF03144; GTP_EFTU_D2; 1.
CC DR Pfam; PF03143; GTP_EFTU_D3; 1.
CC DR PRINTS; PR00315; ELONGATNFCT.
CC DR PROSITE; PS00301; EFACOR_GTP; 1.
CC KW Elongation factor; Protein biosynthesis; GTP-binding.
FT NP_BIND 19 26 GTP (BY SIMILARITY).
FT NP_BIND 81 85 GTP (BY SIMILARITY).
FT NP_BIND 136 139 GTP (BY SIMILARITY).
SQ SEQUENCE 395 AA; 42973 MW; 6FA7C1CF90C72C89 CRC64;

Query Match 4.6%; Score 7; DB 1; Length 395;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 VVGSAL 19
DB 170 VVGSAL 176

RESULT 9
EFTU_CHACO STANDARD: PRT; 408 AA.
AC P50371;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Elongation factor Tu (EF-Tu).
GN TUPA.
OS Chara comitens.
OS Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Characeae; Chara.
OX NCBI_TaxID=13779;

```

```

RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=95392822; PubMed=7663757;
RA      Delwiche C.F., Kuhlse M., Palmer J.D.;
RT      "Phylogenetic analysis of tufa sequences indicates a cyanobacterial
RT      origin of all plastids."
RL      M01. Phylogenetic. Evol. 4:110-128(1995).
CC      -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC      AMINOACYL-tRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC      BIOSYNTHESIS.
CC      -1- SUBCELLULAR LOCATION: Chloroplast.
CC      -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC      EF-TU/EF-1A SUBFAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: U09425; AA87685.1; -
CC      DR      HSSP: P02990; 1EFU.
CC      DR      Mendel: 13588; CHACO; TuFA.1.
CC      DR      InterPro: IPR000795; GTP_EFTU.
CC      DR      InterPro: IPR004161; GTP_EFTU_D2.
CC      DR      InterPro: IPR004160; GTP_EFTU_D3.
CC      DR      InterPro: IPR00009; GTP_EFTU; 1.
CC      DR      Pfam: PF00014; GTP_EFTU_D2; 1.
CC      DR      Pfam: PF03143; GTP_EFTU_D3; 1.
CC      DR      PRINTS: PR00315; ELONGATNFACT.
CC      DR      PROSITE: PS00301; EFACITOR_GTP; 1.
CC      KW      Elongation factor; Protein biosynthesis; Chloroplast;
CC      GTP-binding.
CC      KM      GTP-binding.
CC      FT      NP_BIND 19 26 GTP (BY SIMILARITY).
CC      FT      NP_BIND 81 85 GTP (BY SIMILARITY).
CC      FT      NP_BIND 136 139 GTP (BY SIMILARITY).
CC      FT      FT BIND 408 AA; 45321 MW; DAA452E42B0C3489 CRC64;
CC      SQ      SEQUENCE
CC
Query Match 4.6%; Score 7; DB 1; Length 408;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RT      "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT      VIII."
RL      Science 265:2077-2082(1994).
CC      -1- SIMILARITY: TO C.ELEFANS EEBD8.9 AND S.POMBE SPAC16E8.13.
CC      -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: U11582; CAB34905.1; -
CC      DR      PIR: S46825; S46825.
CC      DR      SGD: S0001002; YHL010C.
CC      DR      InterPro: IPR001841; Znf_ring.
CC      DR      InterPro: IPR001607; Zf-UBP.
CC      DR      Pfam: PF00092; Zf-C3HC4; 1.
CC      DR      Pfam: PF02148; Zf-UBP; 1.
CC      DR      SMART: SM00184; RING; 1.
CC      DR      SMART: SM00290; Znf-UBP; 1.
CC      DR      PROSITE: PS00518; ZF_RING_1; FALSE_NEG.
CC      DR      PROSITE: PS50089; ZF_RING_2; 1.
CC      KW      Hypothetical protein; Zinc-finger.
CC      FT      ZN_RING 240 280 RING-TYPE.
CC      FT      ZN_RING 585 AA; 67503 MW; 88FF670CC73A1263 CRC64;
CC      SQ      SEQUENCE
CC
Query Match 4.6%; Score 7; DB 1; Length 585;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=95392822; PubMed=7663757;
RA      Delwiche C.F., Kuhlse M., Palmer J.D.;
RT      "Phylogenetic analysis of tufa sequences indicates a cyanobacterial
RT      origin of all plastids."
RL      M01. Phylogenetic. Evol. 4:110-128(1995).
CC      -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC      AMINOACYL-tRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC      BIOSYNTHESIS.
CC      -1- SUBCELLULAR LOCATION: Chloroplast.
CC      -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC      EF-TU/EF-1A SUBFAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: U09425; AA87685.1; -
CC      DR      HSSP: P02990; 1EFU.
CC      DR      Mendel: 13588; CHACO; TuFA.1.
CC      DR      InterPro: IPR000795; GTP_EFTU.
CC      DR      InterPro: IPR004161; GTP_EFTU_D2.
CC      DR      InterPro: IPR004160; GTP_EFTU_D3.
CC      DR      InterPro: IPR00009; GTP_EFTU; 1.
CC      DR      Pfam: PF00014; GTP_EFTU_D2; 1.
CC      DR      Pfam: PF03143; GTP_EFTU_D3; 1.
CC      DR      PRINTS: PR00315; ELONGATNFACT.
CC      DR      PROSITE: PS00301; EFACITOR_GTP; 1.
CC      KW      Elongation factor; Protein biosynthesis; Chloroplast;
CC      GTP-binding.
CC      KM      GTP-binding.
CC      FT      NP_BIND 19 26 GTP (BY SIMILARITY).
CC      FT      NP_BIND 81 85 GTP (BY SIMILARITY).
CC      FT      NP_BIND 136 139 GTP (BY SIMILARITY).
CC      FT      FT BIND 408 AA; 45321 MW; DAA452E42B0C3489 CRC64;
CC      SQ      SEQUENCE
CC
Query Match 4.6%; Score 7; DB 1; Length 408;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RT      "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT      VIII."
RL      Science 265:2077-2082(1994).
CC      -1- SIMILARITY: TO C.ELEFANS EEBD8.9 AND S.POMBE SPAC16E8.13.
CC      -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: U11582; CAB34905.1; -
CC      DR      PIR: S46825; S46825.
CC      DR      SGD: S0001002; YHL010C.
CC      DR      InterPro: IPR001841; Znf_ring.
CC      DR      InterPro: IPR001607; Zf-UBP.
CC      DR      Pfam: PF00092; Zf-C3HC4; 1.
CC      DR      Pfam: PF02148; Zf-UBP; 1.
CC      DR      SMART: SM00184; RING; 1.
CC      DR      SMART: SM00290; Znf-UBP; 1.
CC      DR      PROSITE: PS00518; ZF_RING_1; FALSE_NEG.
CC      DR      PROSITE: PS50089; ZF_RING_2; 1.
CC      KW      Hypothetical protein; Zinc-finger.
CC      FT      ZN_RING 240 280 RING-TYPE.
CC      FT      ZN_RING 585 AA; 67503 MW; 88FF670CC73A1263 CRC64;
CC      SQ      SEQUENCE
CC
Query Match 4.6%; Score 7; DB 1; Length 585;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



```

RN [4]
RP SEQUENCE OF 102-587 FROM N.A.
RC TISSUE-Placenta;
RX MEDLINE-96105015; PubMed-8521387;
RA Taylor S.M., Freemantle S.J., Moran R.G.;
RT "Structural organization of the human folypoly-gamma-glutamate
synthetase gene: evidence for a single genomic locus.";
RL Cancer Res. 55:6030-6034(1995).
RN (5)
RP CHARACTERIZATION.
RX MEDLINE-87157665; PubMed-3828320;
RA Cichowicz D.J., Shane B.;
RT "Mammalian folypoly-gamma-glutamate synthetase. 1. Purification and
general properties of the hog liver enzyme.";
RL Biochemistry 26:504-513(1987).
CC -1- FUNCTION: CONVERSION OF FOLATES TO POLYGLUTAMATE DERIVATIVES. THIS
CC ALLOWS TISSUES TO CONCENTRATE FOLATE AT HIGHER LEVELS THAN IN
CC PLASMA.
CC -1- CATALYTIC ACTIVITY: ATP + (tetrahydrofolyl-[glu])(N) + L-glutamate
CC = ADP + phosphate + (tetrahydrofolyl-[glu])(N+1).
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Mitochondrial and cytoplasmic.
CC -1- ALTERNATIVE PRODUCTS: A SINGLE NUCLEAR GENE PRODUCES BOTH FORMS
CC BY USE OF ALTERNATIVE INITIATION CODONS IN THE SAME READING FRAME.
CC -1- SIMILARITY: BELONGS TO THE FOLYLPOLYGLUTAMATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M98045; AAA35852.1; ALT_INIT.
DR EMBL: U14939; AAA85815.1; -.
DR EMBL: U40868; AAA87568.1; -.
DR EMBL: U40863; AAA87568.1; JOINED.
DR EMBL: U40864; AAA87568.1; JOINED.
DR EMBL: U40865; AAA87568.1; JOINED.
DR EMBL: U40866; AAA87568.1; JOINED.
DR EMBL: U40867; AAA87568.1; JOINED.
DR EMBL: U24253; AAC13871.1; -.
DR EMBL: U24252; AAC13871.1; JOINED.
DR PIR: A46281; A46281.
DR HSSP: P15925; 1FGS.
DR MIM: 136510; -.
DR InterPro: IPR001645; FOLYLPOLYGLU_SYNT.
DR InterPro: IPR004101; Mur_Ligase_C.
DR Pfam: PF02875; Mur_Ligase_C.1.
DR PROSITE: PS01011; FOLYLPOLYGLU_SYNT_1; 1.
DR PROSITE: PS01012; FOLYLPOLYGLU_SYNT_2; 1.
KW Ligase; One-carbon metabolism; ATP-binding; Mitochondrion;
KW Transit peptide; Alternative Initiation.
FT TRANSIT 1 42 MITOCHONDRION.
FT CHAIN 1 587 FOLYLPOLYGLUTAMATE SYNTHASE.
FT INIT_MET 43 43 FOR CYTOPLASMIC ISOFORM.
FT NP_BIND 103 109 ATP (POTENTIAL).
FT CONFLICT 22 22 I -> V (IN REF. 3).
FT CONFLICT 22 22 I -> V (IN REF. 3).
SO SEQUENCE 587 AA; 64609 MW; 5AF814095F77E5C CRC64;

Query Match 4.6%; Score 7; DB 1; Length 587;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 54 NAALALQ 60
|||||||
DB 305 NAALALQ 311

RESULT 12
FOLC_MOUSE STANDARD; PRT; 587 AA.
ID FOLC_MOUSE

```

```

AC P48760;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Polypolyglutamate synthase, mitochondrial precursor (EC 6.3.2.17)
DE (Folypoly-gamma-glutamate synthetase) (PF05).
GN PF05.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE-96070787; PubMed-7592937;
RA Roy K., Mitsugi K., Stirling S., Shane B., Strohak F.M.;
RT "Different antifolate-resistant L1210 cell variants with either
RT increased or decreased folypolyglutamate synthetase gene expression
RT at the level of mRNA transcription.";
RL J. Biol. Chem. 270:26918-26922(1995).
RN (2)
RP SEQUENCE FROM N.A.
RX STRAIN=DRA/2;
RX MEDLINE-96180977; PubMed-8605241;
RA Spinella M.J., Bigle K.E., Goldman I.D.;
RT "Molecular cloning of murine folypoly-gamma-glutamate synthetase.";
RL Biochim. Biophys. Acta 1305:11-14(1996).
CC -1- FUNCTION: CONVERSION OF FOLATES TO POLYGLUTAMATE DERIVATIVES. THIS
CC ALLOWS TISSUES TO CONCENTRATE FOLATE AT HIGHER LEVELS THAN IN
CC PLASMA.
CC -1- CATALYTIC ACTIVITY: ATP + (tetrahydrofolyl-[glu])(N) + L-glutamate
CC = ADP + phosphate + (tetrahydrofolyl-[glu])(N+1).
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Mitochondrial and cytoplasmic.
CC -1- ALTERNATIVE PRODUCTS: A SINGLE NUCLEAR GENE PRODUCES BOTH FORMS
CC BY USE OF ALTERNATIVE INITIATION CODONS IN THE SAME READING FRAME.
CC -1- SIMILARITY: BELONGS TO THE FOLYLPOLYGLUTAMATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U33557; AAC52812.1; -.
DR EMBL: U32197; AAC52426.1; -.
DR HSSP: P15925; 1FGS.
DR MGD: MGI:95576; Fpgs.
DR InterPro: IPR001645; FOLYLPOLYGLU_SYNT.
DR Pfam: PF01225; Mur_Ligase_1.
DR PROSITE: PS01011; FOLYLPOLYGLU_SYNT_1; 1.
DR PROSITE: PS01012; FOLYLPOLYGLU_SYNT_2; 1.
KW Ligase; One-carbon metabolism; ATP-binding; Mitochondrion;
KW Transit peptide; Alternative Initiation.
FT TRANSIT 1 42 MITOCHONDRION.
FT CHAIN 1 587 FOLYLPOLYGLUTAMATE SYNTHASE.
FT INIT_MET 43 43 FOR CYTOPLASMIC ISOFORM.
FT NP_BIND 103 109 ATP (POTENTIAL).
FT CONFLICT 27 27 P -> A (IN REF. 2).
FT CONFLICT 35 35 G -> W (IN REF. 2).
FT CONFLICT 75 76 OL -> HV (IN REF. 2).
FT CONFLICT 129 129 R -> S (IN REF. 2).
FT CONFLICT 138 138 D -> E (IN REF. 2).
FT CONFLICT 138 138 G -> S (IN REF. 2).
FT CONFLICT 491 491 G -> S (IN REF. 2).
SO SEQUENCE 587 AA; 64907 MW; B43FB8915EB64E05 CRC64;

Query Match 4.6%; Score 7; DB 1; Length 587;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 54 NAALALQ 60
|||||||

```

DB 305 NALALQ 311

RESULT 13

RNI2_MOUSE STANDARD: PRT: 600 AA.

AC 09WTV7;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE RING finger protein 12 (LIM domain interacting RING finger protein)

GN RNF12 OR RLM.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Bach I.;

RT "Opposing developmental functions of positive and negative coregulators of LIM homeodomain factors."

RL Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.

RN [2]

RP FUNCTION.

RX MEDLINE=99364422; PubMed=10431247;

RA Bach I., Rodriguez-Esteban C., Carrier C., Bhushan A., Krones A., Rose D.W., Glass C.K., Andersen B., Izpisua Belmonte J.C., Rosenfeld M.G.;

RT "RIM inhibits functional activity of LIM homeodomain transcription factors via recruitment of the histone deacetylase complex."

RL Nat. Genet. 22:394-399(1999).

CC -1- FUNCTION: ACTS AS A NEGATIVE CO-REGULATOR FOR LIM HOMEODOMAIN TRANSCRIPTION FACTORS. VIA THE RECRUITMENT OF THE SIN3A/HISTONE DEACETYLASE COMPRESSOR COMPLEX.

CC -1- SUBUNIT: ASSOCIATES WITH LIM/HOMEOBOX FACTORS.

CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL: AF069992; AAC34209.1; -

DR MGD: MGI:1342291; Rnf12.

DR InterPro: IPR001841; Znf_ring.

DR Pfam: PF00097; zf-C3HC4; 1.

DR SMART: SM00184; RING_1; FALSE NEG.

DR PROSITE: PS00518; ZF_RING_1; FALSE NEG.

DR PROSITE: PS0089; ZF_RING_2; 1.

DR TRANSCRIPTION regulation; Zinc-finger.

FW DOMAIN 415 484 SER-RICH.

FT ZN_FING 546 587 RING-TYPE.

FT DOMAIN 447 461 POIX-SER.

SO SEQUENCE 600 AA; 66470 MW; E68293530126E4ID CRC64;

Query Match 4.6%; Score 7; DB 1; Length 600;

Best Local Similarity 100.0%; Pred. No. 32;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 GGNSSG 40

DB 443 GGNSSG 449

RESULT 14

PRH_ARATH STANDARD: PRT: 796 AA.

ID PRH_ARATH

AC P48785;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Pathogenesis-related homeodomain protein (PRH).

GN PRH OR PRHA OR AT4G2940 OR F27B13.180.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Equisetophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV, COLUMBIA;

RX MEDLINE=94312876; PubMed=7913642;

RA Korffage U., Trezzini G.F., Meier I., Hahlbrock K., Somssich I.E.;

RT "Plant homeodomain protein involved in transcriptional regulation of a pathogen defense-related gene."

RL Plant Cell 6:695-708(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CV, COLUMBIA;

RX MEDLINE=98012533; PubMed=9351248;

RA Plesch G., Stoermer K., Tovar Torres J., Walden R., Somssich I.E.;

RT "Developmental and auxin-induced expression of the Arabidopsis pha homeobox gene."

RL Plant J. 12:635-647(1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=CV, COLUMBIA;

RX MEDLINE=20083486; PubMed=10617198;

RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G., Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Treyn N., Harris B., Ansoerge W., Brandt P., Grivell L., Rieger M., Weichselgartner M., de Simone V., Obermaier B., Maché R., Muelier M., Kreis M., Deleney M., Puigdomenech P., Watson M., Schmidheini T., Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I., Vos P., Hohnel J., Zimmermann W., Wedler H., Ridley P., Langham S.A., McCullagh B., Bilham L., Robben J., Van der Schueren J., Grimonprez B., Chuang Y.-J., Vandenbussche F., Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E., Wiltzenger T., Bothe G., Ransperger U., Hilbert H., Braun M., Holzer E., Brandt A., Peters S., van Staveren M., Dirke P., Moeljan P., Klein Lankhorst R., Rose M., Hauf J., Koetter P., Benneker S., Hempel S., Feldpausch M., Lambrecht S., Van den Daele H., De Keyser A., Buyshaert C., Gielen J., Villarroel R., De Clercq R., Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S., Clark L., Doggett J., Hall S., Kay M., Leonard N., McClay K., Mayes R., Pettelt A., Rajandream M.A., Lyne M., Benes V., Rechmann S., Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H., Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzi A., Neumann S., Argilou A., Vitale D., Liguori R., Piravandi E., Massenot O., Outgley F., Clabaud G., Muendlein A., Felber R., Schnabl S., Hiller R., Schmidt W., Lecharry A., Aubourg S., Chefor T., Cooke R., Berger C., Monfort A., Casacuberta E., Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A., Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T., Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bialke C., Fishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S., zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K., Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L., Sekhon M., Murray J., Sheel P., Cordes M., Abu-Threiden J., Stoneking T., Kallick J., Graves T., Harmon G., Edwards J., Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D., Minx P., Bentley D., Fulton L., Miller N., Greco T., Kemp K., Kramer J., Fulton L., Mardis E., Dante M., pepin K., Hillier L., Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D., Du H., Ali J., Berghoff A., Jones K., Dione K., Cotton M., Joshi C., Antonolu B., Zidanic M., Strong C., Sun H., Lamer B., Jordan C., Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R., Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S., Granat S., Shohdy N., Hasegawa A., Hamed A., Lodhi M., Johnson A., Chen E., Marra M., Mortensen R., McCombie W.R.;

RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis

```
RT thaliana.";
RL Nature 402:769-777(1999).
CC -1- FUNCTION: SPECIFICALLY BINDS TO THE FUNGAL ELICITOR-RESPONSIVE
CC DNA ELEMENT, 5'-CTAATGTTA-3', OF THE GENE PR2 PROMOTER.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- INDUCTION: BY PATHOGEN INFECTION.
CC -1- SIMILARITY: BELONGS TO THE PHD-ASSOCIATED FAMILY OF HOMEBOX
CC PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L21991; AAA32843.1; -
DR EMBL; U48864; AAC49836.1; -
DR EMBL; AL050352; CAB43669.1; -
DR EMBL; AL161575; CAB79752.1; -
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR001965; PHD.
DR Pfam; PF00046; homeobox; 1.
DR Pfam; PF00628; PHD; 1.
DR SMART; SMO0389; HOX; 1.
DR SMART; SMO0249; PHD; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Nuclear protein; Transcription regulation;
KW Repeat; Zinc-finger.
FT ZN_FING 188 252 PHD-TYPE.
FT DOMAIN 294 303 ASP-RICH (ACIDIC).
FT DNA_BIND 452 511 HOMEBOX.
FT DOMAIN 605 735 5 X 27 AA TANDEM REPEATS.
FT REPEAT 605 631 1.
FT REPEAT 632 658 2.
FT REPEAT 659 685 3.
FT REPEAT 686 712 4.
FT REPEAT 713 735 5 (INCOMPLETE).
FT DOMAIN 738 759 LEUCINE-ZIPPER.
SQ SEQUENCE 796 AA; 90657 MW; 561509A0CAD6175 CRC64;

Query Match 4.6%; Score 7; DB 1; Length 796;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 NCGNNS 39
Db 309 NCGNNS 315

RESULT 15
HMDH_AGRIP STANDARD; PRT; 833 AA.
ID HMDH_AGRIP
AC 076819;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 3-hydroxy-3-methylglutaryl-coenzyme A reductase (EC 1.1.1.34) (HMG-CoA
DE reductase).
GN HMGR.
OS Agrotis ipsilon (Black cutworm moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Noctuidae; Noctulidae; Noctuidae; Agrotis.
OX NCBI_TaxID=56364;
RN [1]
RP SEQUENCE FROM N.A.
RA Dupontet's L., Rossignol F., Coullaud F.;
RT "Molecular cloning and tissue expression of 3-hydroxy-3-methylglutaryl
RT coenzyme A reductase of the Black cutworm Agrotis ipsilon.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
```

```
CC -1- FUNCTION: SYNTHESIS OF MEVALONATE FOR THE PRODUCTION OF NON-
CC STEROL ISOPRENoids, WHICH ARE ESSENTIAL FOR GROWTH
CC DIFFERENTIATION (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: (R)-mevalonate + CoA + 2 NADP(+) = (S)-3-
CC hydroxy-3-methylglutaryl-CoA + 2 NADPH.
CC -1- ENZYME REGULATION: THE ACTIVITY OF HMG-COA-REDUCTASE IS
CC SUPPRESSED BY EXOGENOUS MEVALONATE.
CC -1- PATHWAY: ISOPRENOID BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
CC reticulum.
CC -1- SIMILARITY: BELONGS TO THE HMG-COA REDUCTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ009675; CA08775.1; -
DR InterPro; IPR002202; HMG-COA_red.
DR InterPro; IPR000731; HMGCR_patched_5TM.
DR Pfam; PF00368; HMG-COA_red; 1.
DR PRINTS; PR00071; HMGCOARDTASE.
DR PROSITE; PS00066; HMG_COA_REDUCTASE_1; 1.
DR PROSITE; PS00318; HMG_COA_REDUCTASE_2; 1.
DR PROSITE; PS01192; HMG_COA_REDUCTASE_3; 1.
DR PROSITE; PS50065; HMG_COA_REDUCTASE_4; 1.
DR PROSITE; PS50156; SSD; 1.
KW Oxidoreductase; Glycoprotein; Endoplasmic reticulum; Transmembrane;
KW Isoprene biosynthesis; NADP.
FT DOMAIN 1 321 MEMBRANE-BOUND.
FT DOMAIN 322 419 LINKER.
FT DOMAIN 420 833 CATALYTIC.
FT TRANSMEM 10 32 POTENTIAL.
FT TRANSMEM 91 117 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
FT TRANSMEM 301 321 POTENTIAL.
FT ACT_SITE 504 504 BY SIMILARITY.
FT ACT_SITE 711 711 BY SIMILARITY.
FT ACT_SITE 809 809 GENERAL BASE (BY SIMILARITY).
FT CARBOHYD 680 680 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 715 715 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 720 720 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 813 813 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 825 825 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 833 AA; 89789 MW; 7E9ADD4B4439A035 CRC64;

Query Match 4.6%; Score 7; DB 1; Length 833;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 SSSALAG 21
Db 682 SSSALAG 688
```

Search completed: October 11, 2002, 21:19:28
Job time : 12 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 11, 2002, 21:15:28 ; Search time 27 Seconds
(without alignments)
967.490 Million cell updates/sec

Title: US-09-543-407-5
Perfect score: 151
Sequence: 1 MKLLKVAFAFAIVSSGALA.....DSSVMVRQVGFNNATANYQ 151

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- 1: SP archaea:*
- 2: SP bacteria:*
- 3: SP fungi:*
- 4: SP human:*
- 5: SP invertebrate:*
- 6: SP mammal:*
- 7: SP mhc:*
- 8: SP organelle:*
- 9: SP phage:*
- 10: SP plant:*
- 11: SP rodent:*
- 12: SP virus:*
- 13: SP vertebrate:*
- 14: SP unclassified:*
- 15: SP virus:*
- 16: SP bacteriap:*
- 17: SP archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127	84.1	152	033802	033802 salmonella
2	72	47.7	76	054069	054069 salmonella
3	13	8.6	19	09R478	09R478 salmonella
4	11	7.3	29	09S3J5	09S3J5 escherichia
5	11	7.3	152	093024	093024 escherichia
6	8	5.3	436	09XA68	09XA68 streptomyce
7	8	5.3	552	09XAR7	09XAR7 streptomyce
8	7	4.6	62	09XA91	09XA91 streptomyce
9	7	4.6	138	09LICO	09XA91 streptomyce
10	7	4.6	145	10 09SDX8	09SDX8 brassica na
11	7	4.6	145	16 099U60	099U60 brassica na
12	7	4.6	206	2 09AE10	09AE10 corynebacte
13	7	4.6	229	16 09CLF5	09CLF5 pasteurella
14	7	4.6	236	2 09AOD9	09AOD9 lactococcus
15	7	4.6	238	11 091X24	091X24 mus musculu
16	7	4.6	267	10 094AN7	094AN7 arabidopsis

17	7	4.6	268	10 09F102	09F102 arabidopsis
18	7	4.6	284	10 042375	042375 arabidopsis
19	7	4.6	284	10 09S335	09S335 arabidopsis
20	7	4.6	303	3 096W96	096W96 aspergillus
21	7	4.6	304	3 092194	092194 aspergillus
22	7	4.6	305	11 09CR05	09CR05 mus musculu
23	7	4.6	311	10 09LNY7	09LNY7 arabidopsis
24	7	4.6	315	2 09L1R5	09L1R5 streptomyce
25	7	4.6	317	10 023220	023220 arabidopsis
26	7	4.6	317	16 09HIS2	09HIS2 rhizobium 1
27	7	4.6	320	2 09KHC9	09KHC9 streptomyce
28	7	4.6	320	2 09K418	09K418 streptomyce
29	7	4.6	323	10 09MA45	09MA45 arabidopsis
30	7	4.6	325	5 09XU26	09XU26 caenorhabdi
31	7	4.6	347	10 09C6X9	09C6X9 arabidopsis
32	7	4.6	351	8 09BAB4	09BAB4 polytoma ob
33	7	4.6	358	10 09FC34	09FC34 arabidopsis
34	7	4.6	367	16 092A84	092A84 listeria in
35	7	4.6	378	5 076736	076736 dictyostell
36	7	4.6	387	16 005313	005313 mycobacteri
37	7	4.6	388	5 044529	044529 caenorhabdi
38	7	4.6	391	11 09D4L4	09D4L4 mus musculu
39	7	4.6	395	4 096M17	096M17 homo sapien
40	7	4.6	395	4 096KC9	096KC9 homo sapien
41	7	4.6	396	6 095JX2	095JX2 macaca fasc
42	7	4.6	396	6 095JW6	095JW6 macaca fasc
43	7	4.6	404	2 09K4Z3	09K4Z3 moritella s
44	7	4.6	404	2 09K4Y8	09K4Y8 moritella s
45	7	4.6	413	2 051691	051691 paracoccus

ALIGNMENTS

RESULT 1									
ID	033802	PRELIMINARY:	PRT:	152 AA.					
AC	033802:								
DT	01-JAN-1998 (TREMBLrel. 05, Created)								
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)								
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)								
DE	AGFA PROTEIN (FRAGMENT).								
GN	AGFA.								
OS	Salmonella typhimurium.								
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;								
CC	Salmonella.								
OX	NCBI_TaxID=602;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=98053981; PubMed=9393832;								
RA	Sukhopolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeiffer J.D.,								
RA	Normark S.J., Rhen M.;								
RT	"Expression of thin, aggregative fimbriae promotes interaction of								
RT	Salmonella typhimurium SR-11 with mouse small intestinal epithelial								
RT	cells.";								
RL	Infect. Immun. 65:5320-5325(1997).								
DR	EMBL: AJ000514; CA04151.1; ..								
FT	NON_TER 152								
SO	SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;								
Query Match 84.1%; Score 127; DB 2; Length 152;									
Best Local Similarity 100.0%; Pred. No. 5.5e-117;									
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	20	AGVPMWGCGNHNGSGPDSLISYOGSANAALALOSDARKSETTTTOSGYGNGA	79						
DB	20	AGVPMWGCGNHNGSGPDSLISYOGSANAALALOSDARKSETTTTOSGYGNGA	79						
QY	80	DVGGGADNSTIELTONGGRNATTIDQNAKNSDITVGGYGNNALVNOTASDSSVMWRQ	139						
DB	80	DVGGGADNSTIELTONGGRNATTIDQNAKNSDITVGGYGNNALVNOTASDSSVMWRQ	139						
QY	140	VGGGNNALVNOTASDSSVMWRQ	146						

Db 140 VGFGNNA 146

RESULT 2

054069 PRELIMINARY; PRT; 76 AA.

DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SEF17 FIMBRIN (FRAGMENT).

OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.

OX NCBI_TaxID=592;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SE30;
RA Cox J.M., Egilezos S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae."
RL Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.

DR EMBL: U53207; AAA98671.1; -.

FT NON_TER 1
FT 76
SQ SEQUENCE 76 AA; 7704 MW; 2FDS411241A7BCB1 CRC64;

Query Match 47.7%; Score 72; DB 2; Length 76;
Best Local Similarity 100.0%; Pred. No. 4e-63;

Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 GGGNSGPDSTLSIYQGSANALALQSDARKSETTITQSGYGCAGDVGQGDNSTIELT 93
Db 5 GGGNSGPDSTLSIYQGSANALALQSDARKSETTITQSGYGCAGDVGQGDNSTIELT 64

Qy 94 ONGFNNATIDQ 105
Db 65 ONGFNNATIDQ 76

RESULT 3

09R4T8 PRELIMINARY; PRT; 19 AA.

DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE TYPE 3 FIMBRIAE (FRAGMENT).

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.

OX NCBI_TaxID=602;

RN [1]

RP SEQUENCE.
RX MEDLINE=95102231; PubMed=7803933;
RA Stolpe H., Grund S., Schroder W.;

RT "Purification and partial characterization of type 3 fimbriae from
RT Salmonella typhimurium var. Copenhagen."
RL Int. J. Med. Microbiol. Virol. Parasitol. Infect. Dis. 281:8-15(1994).

SQ SEQUENCE 19 AA; 1910 MW; 9ADB859D411E551 CRC64;

Query Match 8.6%; Score 13; DB 2; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 GVPFOMGGGGNNH 33
Db 1 GVPFOMGGGGNNH 13

RESULT 4

09S3J5 PRELIMINARY; PRT; 29 AA.

DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CURLIN SUBUNIT MONOMER (FRAGMENT).

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC TRANSPOSON-INSERTION SEQUENCE IS1;
RX MEDLINE=99314153; PubMed=10386375;

RA La Ragione R.M., Collighan R.J., Woodward M.J.;

RT "Non-curlation of Escherichia coli O78:K60 isolates associated with
RT ISI inserti on in cs9b and reduced persistence in poultry infection."
RL FEMS Microbiol. Lett. 175:247-253(1999).

DR EMBL: AJ131756; CAB45380.1; -.

FT NON_TER 29
FT 29
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBE243 CRC64;

Qy 15 SGSALAGVPO 25
Db 15 SGSALAGVPO 25

Qy 15 SGSALAGVPO 25
Db 15 SGSALAGVPO 25

RESULT 5

093U24 PRELIMINARY; PRT; 152 AA.

DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CSGA PROTEIN.

GN CSGA.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=43895 RED VARIANT;
RX MEDLINE=21218556; PubMed=11319125;

RA Uhlich G.A., Keen J.E., Elder R.O.;

RT "Mutations in the cs9b Promoter Associated with Variations in Curl1
RT Expression in Certain Strains of Escherichia coli O157:H7."
RL Appl. Environ. Microbiol. 67:2367-2370(2001).

DR EMBL: AF275733; AAK53212.1; -.

SQ SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;

Qy 15 SGSALAGVPO 25
Db 15 SGSALAGVPO 25

Qy 15 SGSALAGVPO 25
Db 15 SGSALAGVPO 25

RESULT 6

09XA68 PRELIMINARY; PRT; 436 AA.

DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE PUTATIVE TRANSMEMBRANE TRANSPORT PROTEIN.
SCGD3.10C.
GN Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RT Mol. Microbiol. 21:77-96(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AL096822; CAB46932.1;
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr.1.
DR Transmembrane.
SQ SEQUENCE 436 AA; 45991 MW; AFD05F0D59FCA49 CRC64;

Query Match 5.3%; Score 8; DB 2; Length 436;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 GSALAGV 23
|||||||
DB 182 GSALAGV 189

RESULT 7
O9XAR7 PRELIMINARY; PRT; 552 AA.
AC O9XAR7;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NUON, NADH DEHYDROGENASE SUBUNIT.
GN NUON.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL Mol. Microbiol. 21:77-96(1996).
CC -1- CATALYTIC ACTIVITY: NADH + URICUINONE = NAD(+) + URICUINOL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AL078618; CAB44518.1;
DR InterPro: IPR001750; oxidored_q1.
DR Pfam: PF00361; oxidored_q1.1.
DR NAD: Oxidoreductase; Transmembrane.
SQ SEQUENCE 552 AA; 57362 MW; A74E40FEA914ED77 CRC64;

Query Match 5.3%; Score 8; DB 2; Length 552;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 VAAFAIV 13
|||||||
DB 71 VAAFAIV 78

RESULT 8
O9XA91 PRELIMINARY; PRT; 62 AA.
AC O9XA91;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 6.6 KDA PROTEIN.
GN SCF43A.24C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K., Harris D.;
RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RT Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL096837; CAB48911.1;
DR Hypothetical protein.
SQ SEQUENCE 62 AA; 6605 MW; 7F6AF8437BE1B00 CRC64;

Query Match 4.6%; Score 7; DB 2; Length 62;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 55 AALALQS 61
|||||||
DB 39 AALALQS 45

RESULT 9
O9LIC0 PRELIMINARY; PRT; 138 AA.
AC O9LIC0;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PUTATIVE INTEGRAL MEMBRANE PROTEIN.
GN SC11.18.
OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Oliver K., Harris D.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinaishi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL; AL157953; CAB76405.1; -;
 SQ SEQUENCE 138 AA; 14825 MW; 36D3391E13A919EA CRC64;

Query Match 4.6%; Score 7; DB 2; Length 138;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 SALAGV 23
 Db 54 SALAGVV 60

RESULT 10
 Q9SDX8 PRELIMINARY; PRT; 145 AA.
 AC Q9SDX8;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PHOSPHOENOLPYRUVATE CARBOXYLASE KINASE (FRAGMENT).
 OS Brassica napus (rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustoids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20040078; PubMed=10571893;
 RA Hartwell J., Gill A., Nimmo G.A., Wilkins M.B., Jenkins G.I.,
 RA Nimmo H.G.;
 RT "phosphoenolpyruvate carboxylase kinase is a novel protein kinase
 RT regulated at the level of expression.";
 RL Plant J. 20:333-342(1999).
 CC -1 SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL; AF203482; AAF19404.1; -;
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002290; Ser_thr_Pkinase.
 DR Pfam: PF00069; Pkinase; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; Kinase; Pyruvate; Serine/threonine-protein kinase;
 KW Transferrase.
 RT NON_TER 1 1
 FT NON_TER 145 145
 SQ SEQUENCE 145 AA; 16023 MW; A90350EA032457F1 CRC64;

Query Match 4.6%; Score 7; DB 10; Length 145;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 DSTLSIY 48

Db 46 DSTLSIY 52

RESULT 11
 Q99U60 PRELIMINARY; PRT; 145 AA.
 AC Q99U60;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN SA1261 (HYPOTHETICAL PROTEIN SAV1428).
 GN SA1261 OR SAV1428.
 OS Staphylococcus aureus (strain N315), and
 OS Staphylococcus aureus (strain Mu50).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 OX NCBI_TaxID=158879, 158878;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);
 RX MEDLINE=21311952; PubMed=11418146;
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
 RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekimizu K., Hiraoka H., Kuhara S., Goto S., Yabuuchi J.,
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus
 RT aureus.";
 RL Lancet 357:1225-1240(2001).
 DR EMBL; AP003133; BAB42521.1; -;
 DR EMBL; AP003362; BAB57590.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 145 AA; 16014 MW; 245ECA4104A63875 CRC64;

Query Match 4.6%; Score 7; DB 16; Length 145;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 91 ELTONGF 97
 Db 18 ELTONGF 24

RESULT 12
 Q9AE10 PRELIMINARY; PRT; 206 AA.
 AC Q9AE10;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HOLLIDAY JUNCTION DNA-HELICASE RUVA.
 GN RUVA.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
 OC Corynebacterium.
 OX NCBI_TaxID=1718;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC13032;
 RX MEDLINE=98361031; PubMed=9695918;
 RA Wehner L., Schafer A., Burkovski A., Kramer R., Mechold U.,
 RA Maile H., Puhler A., Kalinowski J.;
 RT "The role of the Corynebacterium glutamicum *rel* gene in (p)ppGpp
 RT metabolism.";
 RL Microbiology 144:1853-1862(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC13032;
 RX Berens S., Kalinowski J., Puhler A.;

RT "The role of *Corynebacterium glutamicum* secretion genes *secD*, *secE* and
RT *secG* in transporting the Streptomyces griseus alpha-amylase."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF038651; AAK19839.1; -
DR HSP; P40832; IBVS.
DR InterPro: IPR003583; HHM_1.
DR InterPro: IPR000085; RUVA.
DR Pfam: PF01330; RUVA.1.
DR Pfam: PF02904; RUVA.1.1.
DR ProDom: PD006268; RUVA.1.
DR SMART: SM00278; Hnh1; 1.
KW Helicase.
SQ SEQUENCE 206 AA; 21476 MW; F8F0C96358C34E8B CRC64;
Query Match 4.6%; Score 7; DB 2; Length 206;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 KVAFAA 11
DB 130 KVAFAA 136
RESULT 13
Q9CLF5 PRELIMINARY; PRT; 229 AA.
AC Q9CLF5;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE DOD.
DE DOD_1 OR PM1280.
GN *Pasteurella multocida*.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PM70;
RX MEDLINE-2145866; PubMed-11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of *Pasteurella multocida* pm70."
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL: AE006167; AAK03364.1; -
DR InterPro: IPR000056; Ribul_P_3-epim.
DR Pfam: PF00834; Ribul_P_3-epim; 1.
KW Complete proteome.
SQ SEQUENCE 229 AA; 25461 MW; 61D123691A499BE7 CRC64;
Query Match 4.6%; Score 7; DB 16; Length 229;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 13 VVGSAL 19
DB 204 VVGSAL 210
RESULT 14
Q9AOD9 PRELIMINARY; PRT; 236 AA.
AC Q9AOD9;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE LRRB.
GN RRB.
OS *Lactococcus lactis*.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC *Lactococcus*.
OX NCBI_TaxID=1358;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-MG1363; PubMed-10784052;
RX MEDLINE-20244638;
RA O'Connell-Motherway M., van Sinderen D., Morel-Deville F.,
RA Fitzgerald G.F., Ehrlich S.D., Morel P.;
RT "Six putative two-component regulatory systems isolated from
RT *Lactococcus lactis* subsp. cremoris MG1363."
RL Microbiology 146:935-947(2000).
CC -1 SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
CC REGULATORY COMPONENTS OF SENSOR TRANSDUCTION SYSTEMS.
CC -1 SIMILARITY: TO OTHER BACTERIAL REGULATORY PROTEINS INVOLVED IN
CC SIGNAL TRANSDUCTION.
DR EMBL: AF176556; AAG53726.1; -
DR HSP; P08402; IB00.
DR InterPro: IPR001789; Response-reg.
DR InterPro: IPR001867; Trans-reg.C.
DR Pfam: PF00072; Response-reg; 1.
DR Pfam: PF00486; Trans-reg.C; 1.
DR SMART: SM00448; REC; 1.
KW DNA-binding; Phosphorylation; Sensory transduction;
KW Transcription regulation.
SQ SEQUENCE 236 AA; 27273 MW; 3DD5808FA45D0C4E CRC64;
Query Match 4.6%; Score 7; DB 2; Length 236;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 127 NOTASDS 133
DB 124 NOTASDS 130
RESULT 15
Q91X24 PRELIMINARY; PRT; 238 AA.
AC Q91X24;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE SIMILAR TO INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-COLON;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC012723; AAH12723.1; -
SQ SEQUENCE 238 AA; 25346 MW; DDBAFD5DB0413CF7 CRC64;
Query Match 4.6%; Score 7; DB 11; Length 238;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 15 SGSALAG 21
DB 23 SGSALAG 29
Search completed: October 11, 2002, 21:20:09
Job time : 29 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 11, 2002, 21:03:23 ; Search time 32 seconds
(without alignments)
524.129 Million cell updates/sec

Title: US-09-543-407-5

Perfect score: 773

Sequence: 1 MKLTKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	773	100.0	151	AA874625	Agfa sequence, Sa
2	773	100.0	151	AA836341	Salmonella enterit
3	768	99.4	151	AAW23570	Salmonella enterit
4	695	89.9	151	AA836346	Agfa::PT3#1 amino
5	693	89.7	151	AA836347	Agfa::PT3#2 amino
6	692	89.5	151	AA836353	Agfa::PT3#8 amino
7	691	89.4	151	AA836349	Agfa::PT3#4 amino
8	690	89.3	151	AA836352	Agfa::PT3#7 amino
9	689	89.1	151	AA836350	Agfa::PT3#5 amino
10	683	88.4	151	AA836354	Agfa::PT3#9 amino
11	682	88.2	151	AA836351	Agfa::PT3#6 amino

12	681	88.1	151	21	AA836355	Agfa::PT3#10 amino
13	658	85.1	151	21	AA836348	Agfa::PT3#3 amino
14	588	76.1	120	15	AA862761	Agfa sequence, Sa
15	588	76.1	120	18	AAW23569	Salmonella enterit
16	586	75.8	151	21	AA836343	Escherichia coli C
17	508	65.7	142	15	AA852664	Fibronectin bind P
18	436	56.4	122	15	AA852663	Salmonella enterit
19	237	30.7	45	21	AA836316	Salmonella enterit
20	132	17.1	22	21	AA836318	Salmonella enterit
21	123	15.9	23	21	AA836321	Salmonella enterit
22	123	15.9	23	21	AA836326	Salmonella enterit
23	123	15.9	23	21	AA836338	Salmonella enterit
24	115	14.9	22	21	AA836320	Salmonella enterit
25	115	14.9	22	21	AA836325	Salmonella enterit
26	115	14.9	22	21	AA836339	Salmonella enterit
27	112.5	14.6	151	21	AA836344	Escherichia coli C
28	111.5	14.4	151	21	AA836342	Salmonella enterit
29	111	14.4	22	21	AA836322	Salmonella enterit
30	111	14.4	22	21	AA836337	Salmonella enterit
31	111	14.4	22	21	AA836337	Salmonella enterit
32	109	14.1	23	21	AA836337	Salmonella enterit
33	109	14.1	23	21	AA836324	Salmonella enterit
34	109	14.1	23	21	AA836340	Salmonella enterit
35	105	13.6	597	22	AAU08231	Polyptide encode
36	98.5	12.7	850	22	AB865764	Drosophila melanog
37	98.5	12.7	1028	22	AB862708	Drosophila melanog
38	98	12.7	718	21	AA878364	H. pylori bacteria
39	97.5	12.6	738	19	AAW56153	New DNA sequence i
40	96	12.4	19	21	AA836323	Salmonella enterit
41	96	12.4	19	21	AA836328	Salmonella enterit
42	96	12.4	19	21	AA836336	Salmonella enterit
43	94	12.2	892	19	AAW68203	M. catarrhalis str
44	93.5	12.1	287	22	AB865343	Drosophila melanog
45	93.5	12.1	447	21	AA829728	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AA874625 standard; Protein: 151 AA.

AA874625:

26-JUN-1995 (first entry)

Agfa sequence.

Salmonella: Agfa: vaccine.

Salmonella.

MO9425598-A.

10-NOV-1994.

26-APR-1994; 94WO-IB00207.

26-APR-1993; 93US-0054452.

(KING/) KING J.

(UYVT-) UNIT VICTORIA INNOVATION & DEV CORP.

Clouthier SC, Collinson SK, Doran JL, Kay WM;

WPI: 1994-358275/44.

N-PSDB; AAQ87467.

Eliciting an immune response to Salmonella - using attenuated Salmonella strains, vector constructs, or compsns. contg. fibrillar type proteins.

PS Disclosure; Fig7B; 95pp; English.

XX The Salmomella Agfa protein and DNA are used in vaccine and

CC genetic immunization compositions, respectively, to elicit an

CC immune response to Salmomella in animals (e.g. food producing

CC animals) and humans.

XX Sequence 151 AA;

SQ

Query Match 100.0%; Score 773; DB 15; Length 151;

Best Local Similarity 100.0%; Pred. No. 1.8e-65;

Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLKVAFAAIVVSGSALAGVPMQWGGGNNHNGGSSGPDSTLSTIYQGSANAALALQ 60

DB 1 MKLKVAFAAIVVSGSALAGVPMQWGGGNNHNGGSSGPDSTLSTIYQGSANAALALQ 60

QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151

DB 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151

RESULT 2

AAB36341

ID AAB36341 standard; Protein: 151 AA.

XX AAB36341;

AC 26-FEB-2001 (first entry)

XX

DE Salmomella enteritidis Agfa amino acid sequence SEQ ID NO:5.

XX

KM Salmomella: agfa; chromosomal gene replacement; fimbria; epitope;

KV vaccine; immune response; immunogen.

XX

OS Salmomella enteritidis.

XX

PN WO200060102-A2.

XX

PD 12-OCT-2000.

XX

XX 05-APR-2000; 2000WO-CA00356.

XX

XX 05-APR-1999; 990S-0127888.

XX

XX (UYVI-) UNIV VICTORIA.

XX

PI White AP, Doran JL, Collison SK, Kay WM;

XX

DR WPI; 2000-672631/65.

XX

DR N-PSDB; AAC64617.

XX

PT Recombinant agfa gene having a segment replaced by a foreign DNA

PT sequence which encodes foreign epitope or antigen, expresses

PT recombinant Agfa protein useful for eliciting immune response in animal

PT

PS Disclosure; Page 135; 139pp; English.

XX

XX The present invention describes a recombinant agfa gene (I) where a

CC segment of the gene has been replaced by a segment of a foreign DNA

CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended

CC assembly system of strains of Salmomella, Escherichia coli and

CC Enterobacteriaceae for the production of fimbriae comprising recombinant

CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene

CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,

CC comprising separating an amino acid polymer comprising a recombinant

CC Agfa protein containing a replacement segment or segments of foreign

CC amino acid sequence or sequences grown on a Salmomella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the

CC polymer into the animal in conjunction with a carrier or diluent. (1) is

CC useful for the expression of recombinant Agfa protein which is useful

CC for eliciting an immune response in an animal. In a fimbrial presentation

CC system the heterologous antigens are presented in high numbers (up to

CC 500,000 copies/cell) the hybrid fimbria protein possesses both the

CC immunogenicity and adhesion properties relevant for an efficient live

CC vaccine, the carrier fimbrial subunit proteins are usually strong

CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fimbriae are easy and

CC inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention.

XX Sequence 151 AA;

SQ

Query Match 100.0%; Score 773; DB 21; Length 151;

Best Local Similarity 100.0%; Pred. No. 1.8e-65;

Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLKVAFAAIVVSGSALAGVPMQWGGGNNHNGGSSGPDSTLSTIYQGSANAALALQ 60

DB 1 MKLKVAFAAIVVSGSALAGVPMQWGGGNNHNGGSSGPDSTLSTIYQGSANAALALQ 60

QY 61 SDARKSETTITQSGYGNAGADVQGGADNSTIELTONGFRNNATIDQWNAKNSDITVGYG 120

DB 61 SDARKSETTITQSGYGNAGADVQGGADNSTIELTONGFRNNATIDQWNAKNSDITVGYG 120

QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151

DB 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151

RESULT 3

AAW23570

ID AAW23570 standard; Protein: 151 AA.

XX AAW23570;

AC 29-SEP-1997 (first entry)

XX

DE Salmomella enteritidis 27655-3b agfa.

XX

XX Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.

XX

OS Salmomella enteritidis.

XX

XX Key Location/Qualifiers

FT Misc-difference 123

FT /note= "Encoded by GCC"

XX

XX US5635617-A.

XX

XX 03-JUN-1997.

XX

XX 26-APR-1993; 930S-0054452.

XX

XX 26-APR-1994; 940S-0233788.

XX

XX 26-APR-1993; 930S-0054452.

XX

XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.

XX

XX Clouthier SC, Collinson SK, Doran JL, Kay WM;

XX

XX WPI; 1997-309886/28.

XX

XX N-PSDB; AAT74142.

XX

XX Isolated Salmomella gene agfa - used for diagnosis of Salmomella or

PT enteropathogenic bacteria of the Enterobacteria family

DR WP1: 2000-672631/65.
 XX N-PSDB: AAC64623.
 PR Recombinant agfa gene having a segment replaced by a foreign DNA
 PT sequence which encodes foreign epitope or antigen, expresses
 PT recombinant Agfa protein useful for eliciting immune response in animal
 XX
 PS Disclosure: Page 136; 139pp; English.
 CC
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CSga and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant
 CC Agfa protein containing a replacement segment or segments of foreign
 CC amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful
 CC for eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention.
 XX
 XX Sequence 151 AA:
 SQ
 Query Match 89.7%; Score 693; DB 21; Length 151;
 Best Local Similarity 90.7%; Pred. No. 6.5e-58;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAIYVSGSALAGVVPQMGGGGNHNGGSSGPPSTLSIYOGSANAALALQ 60
 DB 1 MKLLKVAFAIYVSGSALAGVVPQMGGGGNHNGGSSGPPSTLSIYOGSANAALALQ 60
 QY 61 SDARKSETTTQSGYNGADVGGADNSTIELTQNGFRNNATIDQNNANNSDITYGQYGG 120
 DB 61 SDARKSETTTQSGYNGADVGGADNSTIELTQNGFRNNATIDQNNANNSDITYGQYDQ 120
 QY 121 NNAALVNOTASDSSVMYRQYFGNNATANY 151
 DB 121 LVTRVVTHEMAHASVMYRQYFGNNATANY 151
 RESULT 6
 AAB36353
 ID AAB36353 standard; Protein; 151 AA.
 XX AAB36353:
 AC 26-FEB-2001 (first entry)
 DT
 XX Agfa::PT3#8 amino acid sequence SEQ ID NO:26.
 DE
 XX Salmonella; agfa: chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS - synthetic.
 XX

PN WO20060102-A2.
XX
PD 12-OCT-2000.
XX PF 05-APR-2000; 2000WO-CA00356.
XX PR 05-APR-1999; 99US-0127888.
XX PA (UYVI-) UNIV VICTORIA.
PI White AP, Doran JL, Collison SK, Kay WM;
DR MPI; 2000-672631/65.
XX N-PADB; AAC64629.

Pt Recombinant agfa gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant Agfa protein useful for eliciting immune response in animal

-
PS Disclosure; Page 138; 139pp; English.

CC The present invention describes a recombinant agfa gene (I) where a CC segment of the gene has been replaced by a segment of a foreign DNA CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (Sefr/Taf) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CagA and AgfA-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (II) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation CC system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrial protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention.

XX Sequence 151 AA:

SQ

Query Match 89.5%; Score 692; DB 21; Length 151;

Best Local Similarity 90.7%; Pred. No. 8, le=58;

Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0.

Dy 1 MKLLKVAFAAIYVSGSALAGVPWPWGGGNGHNGCGSSGPSDSLISYOGSANAAIAIQ 60
|||||
Dd 1 MKLLKAFAAIYVSSGSALAGVPPWGCGGNHNHGCGSSGPSDSLISYOGSANAAIALQ 60
||| : |||
Qy 61 SDARKSETTIGSTGYGCADVGCGANDSTIELTONGRNNAATIDONAKASDIITVOYG 120
LVTFTVVTHEMAHAYCAGADVGOGADNSTIELTONGFRNNATTIDOWNAKNSDITTVOYG 120

Qy 121 NNAAVLVNQTASSSVMYROYGFENNATANDY 151
|||||
Dd 121 NNAAVLVNQTASSSVMYROYGFENNAITAN DY 151

RESULT 7
AADB36349
X X AAB36349 standard; Protein: 151 AA.
AC AAB36349;

SQ Sequence 151 AA;

Query Match 89.3%; Score 690; DB 21; Length 151;
 Best Local Similarity 91.4%; Pred. No. 1.2e-57;
 Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLKVAFAFAIYVSGSALAGVYPQWGGGNGHNGGSSGPDSTLSIYOGSANAALALQ 60
 DB 1 MKLKVAFAFAIYVSGSALAGVYPQWGGGNGHNGGSSGPDSTLSIYOGSANAALALQ 60

QY 61 SDARKSETTITGSGYNGADVGOGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGC 120
 DB 61 SDARKSETTITGSGYNGADVGOGADNSTIELTQNGFRNNATIDQWNAKNSDITVTRVLT 120

QY 121 NNAALVNOTASDSSVWVRQVFGNNATANQY 151
 DB 121 HEMAHANOTASDSSVWVRQVFGNNATANQY 151

RESULT 9
 AAB36350
 ID AAB36350 standard; Protein; 151 AA.
 AC AAB36350;
 XX 26-FEB-2001 (first entry)
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT3#5 amino acid sequence SEQ ID NO:20.
 KW Agfa::PT3#5 amino acid sequence SEQ ID NO:20.
 KM Salmomella; agfa; chromosomal gene replacement; fimbrin; epitope;
 OS vaccine; immune response; immunogen.
 OS Salmomella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 OS WO200060102-AZ.
 PN 12-OCT-2000.
 PD 12-OCT-2000.
 PF 05-APR-2000; 2000WO-CA00356.
 XX 05-APR-1999; 99US-0127888.
 PR 05-APR-1999; 99US-0127888.
 PA (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay MW;
 PI WPI: 2000-672631/65.
 XX N-PSDB: AAC64626.
 DR N-PSDB: AAC64626.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA
 PT sequence which encodes foreign epitope or antigen, expresses
 PT recombinant Agfa protein useful for eliciting immune response in animal
 PT
 PT
 PT
 PS Disclosure; Page 137; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of this aggregative fimbriae (SEF17/TAI) nucleation depended
 CC assembly system of strains of Salmomella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant
 CC Agfa protein containing a replacement segment or segments of foreign
 CC amino acid sequence or sequences grown on a Salmomella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the

CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful
 CC for eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell) the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention.

SQ Sequence 151 AA;

Query Match 89.1%; Score 689; DB 21; Length 151;
 Best Local Similarity 90.7%; Pred. No. 1.6e-57;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLKVAFAFAIYVSGSALAGVYPQWGGGNGHNGGSSGPDSTLSIYOGSANAALALQ 60
 DB 1 MKLKVAFAFAIYVSGSALAGVYPQWGGGNGHNGGSSGPDSTLSIYOGSANAALALQ 60

QY 61 SDARKSETTITGSGYNGADVGOGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGC 120
 DB 61 SDARKSETTITGSGYNGADVGOGADNSTIELTQNGFRNNATIDQWNAKNSDITVTRVLT 120

QY 121 NNAALVNOTASDSSVWVRQVFGNNATANQY 151
 DB 121 NNAALVNOTASDSSVWVRQVFGNNATANQY 151

RESULT 10
 AAB36354
 ID AAB36354 standard; Protein; 151 AA.
 AC AAB36354;
 XX 26-FEB-2001 (first entry)
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT3#9 amino acid sequence SEQ ID NO:28.
 KW Agfa::PT3#9 amino acid sequence SEQ ID NO:28.
 KM Salmomella; agfa; chromosomal gene replacement; fimbrin; epitope;
 OS vaccine; immune response; immunogen.
 OS Salmomella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 OS WO200060102-AZ.
 PN 12-OCT-2000.
 PD 12-OCT-2000.
 PF 05-APR-2000; 2000WO-CA00356.
 XX 05-APR-1999; 99US-0127888.
 PR 05-APR-1999; 99US-0127888.
 PA (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay MW;
 PI WPI: 2000-672631/65.
 XX N-PSDB: AAC64630.
 DR N-PSDB: AAC64630.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA
 PT sequence which encodes foreign epitope or antigen, expresses
 PT recombinant Agfa protein useful for eliciting immune response in animal
 PT
 PT
 PT
 PS Disclosure; Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant
CC Agfa protein containing a replacement segment or segments of foreign
CC amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful
CC for eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention.

Sequence 151 AA:

Query Match 88.4%; Score 683; DB 21; Length 151;

Best Local Similarity 90.1%; Pred. No. 5.7e-57;

Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

OY 1 MKLVAAFAAIVSGSALAGVVPQMGCGGNHNGGSSGPDSTLSIYOGSANAALAQ 60
DB 1 MKLVAAFAAIVSGSALAGVVPQMGCGGNHNGGSSGPDSTLSIYOGSANAALAQ 60
OY 61 SPARKSETTITGSGYGNADVGGADNSTIELTQNGFRNNATTIDOMAKNSDITVGQYGG 120
DB 61 SPARKSETTITGSGYGNADVGGADNSTIELTQNGFRNNATTIDOMAKNSDITVGQYGG 120
OY 121 NNAALVNOTASDSVWVROYGFGNNATANQY 151
DB 121 NNAALVNOTASDSVWVROYGFGNNATANQY 151

RESULT 11
AAB36351
ID AAB36351 standard; Protein: 151 AA.

XX AAB36351;

DT 26-FEB-2001 (first entry)

DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.

KM *Salmonella*: agfa: chromosomal gene replacement; fimbrial; epitope;
KM vaccine; immune response; Immunogen.

OS *Salmonella enteritidis*.

OS *Escherichia coli*.

OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000MO-CA00356.

PR 05-APR-1999; 990US-0127888.

PA (UVVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay WW;

DR WPI: 2000-672631/65.

DR N-PSDB: AAC64627.

XX Recombinant agfa gene having a segment replaced by a foreign DNA

PT sequence which encodes foreign epitope or antigen, expresses

PT recombinant Agfa protein useful for eliciting immune response in animal

PS Disclosure: Page 137: 139pp; English.

CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant
CC Agfa protein containing a replacement segment or segments of foreign
CC amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful
CC for eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention.

Sequence 151 AA:

Query Match 88.2%; Score 682; DB 21; Length 151;

Best Local Similarity 90.7%; Pred. No. 7.1e-57;

Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

OY 1 MKLVAAFAAIVSGSALAGVVPQMGCGGNHNGGSSGPDSTLSIYOGSANAALAQ 60
DB 1 MKLVAAFAAIVSGSALAGVVPQMGCGGNHNGGSSGPDSTLSIYOGSANAALAQ 60
OY 61 SPARKSETTITGSGYGNADVGGADNSTIELTQNGFRNNATTIDOMAKNSDITVGQYGG 120
DB 61 SPARKSETTITGSGYGNADVGGADNSTIELTQNGFRNNATTIDOMAKNSDITVGQYGG 120
OY 121 NNAALVNOTASDSVWVROYGFGNNATANQY 151
DB 121 NNAALVNOTASDSVWVROYGFGNNATANQY 151

RESULT 12
AAB36355
ID AAB36355 standard; Protein: 151 AA.

XX AAB36355;

DT 26-FEB-2001 (first entry)

DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.

KM *Salmonella*: agfa: chromosomal gene replacement; fimbrial; epitope;
KM vaccine; immune response; Immunogen.

OS *Salmonella enteritidis*.

OS *Escherichia coli*.

OS Synthetic.

PN WO200060102-A2.

OY 121 NNAALVNOTASDSSVMROYGFGNNATANDY 151
|
DB 121 NNAALVNOTASDSSVMROYGFGNNATANDY 151

RESULT 14
AAR62761

ID AAR62761 standard; Protein: 120 AA.

AC AAR62761;

DT 26-JUN-1995 (first entry)

DE Agfa sequence.

KM Salmonella; Agfa: vaccine.

OS Salmonella enteritidis 27655-3b.

PN MO9425598-A.

PD 10-NOV-1994.

PE 26-APR-1994; 94MO-IB00207.

PR 26-APR-1993; 93US-0054452.

PA (KING/) KING J.
(UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.

PI Clouthier SC, Collinson SK, Doran JL, Kay WM;

DR WPI: 1994-358275/44.
N-PSDB: AAO73066.

PT Eliciting an immune response to Salmonella - using attenuated

PT Salmonella strains, vector constructs, or compsns. contg.

PT fimbrial type proteins.

PS Disclosure: Fig7A; 95pp; English.

CC The sequence represents the Salmonella enteritidis 27655-3b

CC TnpH mutant strain Agfa protein. The encoding DNA and

CC isolated Agfa protein are used in genetic immunization and vaccine

CC compositions, respectively, to elicit an immune response to

CC Salmonella in animals (e.g. food producing animals) and humans.

SQ Sequence 120 AA;

Query Match 76.1%; Score 588; DB 15; Length 120;

Best Local Similarity 100.0%; Pred. No. 4.1e-48;

Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 VVPQMGCGNHNHGGNSGPDSTLSIYQGSANALALQSDARKSETTITQSGYNGADV 81
|
DB 1 VVPQMGCGNHNHGGNSGPDSTLSIYQGSANALALQSDARKSETTITQSGYNGADV 60

OY 82 GCGADNSTIELTONGFRNNATIDOMNAKNSDITVGOYGNNALVNOTASDS 133
|
DB 61 GCGADNSTIELTONGFRNNATIDOMNAKNSDITVGOYGNNALVNOTASDS 112

RESULT 15
AAM23569

ID AAM23569 standard; Protein: 120 AA.

AC AAM23569;

DT 29-SEP-1997 (first entry)

DE Salmonella enteritidis 27655-3b TnpH mutant agfa fragment.

KM Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.
OS Salmonella enteritidis.

PN US5635617-A.

PD 03-JUN-1997.

PE 26-APR-1993; 93US-0054452.

PR 26-APR-1994; 94US-0233788.

PR 26-APR-1993; 93US-0054452.

PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.

PI Clouthier SC, Collinson SK, Doran JL, Kay WM;

DR WPI: 1997-309886/28.

DR N-PSDB: AAT74141.

PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or

PT enteropathogenic bacteria of the Enterobacteria family

PS Example 2; Fig 7; 85pp; English.

CC The present sequence represents an agfa fragment encoded by an agfa gene

CC fragment derived from Salmonella enteritidis 27655-3b TnpH mutant

CC strain. The nucleic acid can be used to provide diagnostic assays for

CC Salmonella and/or enteropathogenic bacteria of the family

CC Enterobacteria. It can also be used to provide proteins and antibodies

CC which can be used for assays. The nucleic acid sequence can be used to

CC provide probes or primers which can specifically hybridise to nucleic

CC acid molecules from greater than 99% of Salmonella strains that are

CC pathogenic to warm-blooded animals relative to nucleic acid molecules

CC from virtually all other microbial organisms.

SQ Sequence 120 AA;

Query Match 76.1%; Score 588; DB 18; Length 120;

Best Local Similarity 100.0%; Pred. No. 4.1e-48;

Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 VVPQMGCGNHNHGGNSGPDSTLSIYQGSANALALQSDARKSETTITQSGYNGADV 81
|
DB 1 VVPQMGCGNHNHGGNSGPDSTLSIYQGSANALALQSDARKSETTITQSGYNGADV 60

OY 82 GCGADNSTIELTONGFRNNATIDOMNAKNSDITVGOYGNNALVNOTASDS 133
|
DB 61 GCGADNSTIELTONGFRNNATIDOMNAKNSDITVGOYGNNALVNOTASDS 112

Search completed: October 11, 2002, 21:13:17
Job time : 33 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 11, 2002, 21:12:43 : Search time 17 seconds
(without alignments)
216.957 Million cell updates/sec

Title: US-09-543-407-5
Perfect score: 773
Sequence: 1 MKLKVAFAAIVSGSALA.....DSSVMYROYGFGNNATANCY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 100%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2.6/prodata/2/1aa/5A.COMB.pep:*
2: /cgn2.6/prodata/2/1aa/5B.COMB.pep:*
3: /cgn2.6/prodata/2/1aa/6A.COMB.pep:*
4: /cgn2.6/prodata/2/1aa/6B.COMB.pep:*
5: /cgn2.6/prodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2.6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query length	DB ID	Description
1	768	99.4	151	1 US-08-233-788A-59	Sequence 59, Appl
2	588	76.1	120	1 US-08-233-788A-57	Sequence 57, Appl
3	97.5	12.6	738	3 US-08-864-038A-3	Sequence 3, Appl
4	94	12.2	892	4 US-09-336-447A-5	Sequence 5, Appl
5	92	11.9	673	4 US-09-196-387-8	Sequence 8, Appl
6	92	11.9	949	4 US-09-196-387-10	Sequence 10, Appl
7	92	11.9	1327	4 US-09-196-387-2	Sequence 2, Appl
8	89.5	11.6	873	4 US-09-336-447A-13	Sequence 13, Appl
9	89	11.5	943	4 US-09-056-556-204	Sequence 204, Appl
10	81.5	10.5	941	4 US-09-336-447A-9	Sequence 9, Appl
11	81	10.5	1160	3 US-08-808-599A-24	Sequence 24, Appl
12	81	10.5	1864	2 US-08-804-227C-3	Sequence 3, Appl
13	80.5	10.4	437	2 US-08-737-716-2	Sequence 2, Appl
14	80.5	10.4	518	3 US-09-043-123-2	Sequence 2, Appl
15	79	10.2	251	1 US-08-209-747-8	Sequence 8, Appl
16	79	10.2	251	1 US-08-458-298-8	Sequence 8, Appl
17	78.5	10.2	415	4 US-09-025-7698-280	Sequence 280, App
18	78.5	10.2	720	1 US-07-731-157A-2	Sequence 2, Appl
19	78.5	10.2	720	2 US-08-541-780-2	Sequence 2, Appl
20	78	10.1	624	4 US-09-336-447A-7	Sequence 7, Appl
21	78	10.1	906	1 US-08-254-573-2	Sequence 2, Appl
22	78	10.1	906	1 US-08-687-379-2	Sequence 2, Appl
23	78	10.1	906	1 US-08-687-379-2	Sequence 2, Appl
24	78	10.1	906	4 US-08-172-332-1	Sequence 4, Appl
25	78	10.1	2123	4 US-08-968-685A-10	Sequence 1, Appl
26	77.5	10.0	1536	1 US-08-038-682-2	Sequence 2, Appl
27	77.5	10.0	1536	1 US-08-302-832-2	Sequence 2, Appl

28	77.5	10.0	1536	2 US-08-530-198-2	Sequence 2, Appl
29	77.5	10.0	1536	2 US-08-469-880-2	Sequence 2, Appl
30	77.5	10.0	1536	2 US-08-728-470-2	Sequence 2, Appl
31	77.5	10.0	1536	2 US-08-617-697-2	Sequence 2, Appl
32	77.5	10.0	1536	4 US-08-719-641-2	Sequence 2, Appl
33	77.5	10.0	1612	4 US-08-169-927-2	Sequence 2, Appl
34	77	10.0	1026	1 US-08-194-290-7	Sequence 7, Appl
35	77	10.0	1026	2 US-08-614-377A-7	Sequence 7, Appl
36	77	10.0	1026	4 US-09-142-648B-7	Sequence 7, Appl
37	76.5	9.9	211	1 US-08-276-852-34	Sequence 34, Appl
38	76.5	9.9	211	1 US-08-133-011-16	Sequence 16, Appl
39	76.5	9.9	211	1 US-08-322-730A-16	Sequence 16, Appl
40	76.5	9.9	211	1 US-08-387-874-16	Sequence 16, Appl
41	76.5	9.9	211	1 US-08-899-575-34	Sequence 34, Appl
42	76.5	9.9	211	1 US-08-899-575-34	Sequence 34, Appl
43	76.5	9.9	211	2 US-08-383-619-16	Sequence 16, Appl
44	76.5	9.9	211	4 US-08-907-739-16	Sequence 16, Appl
45	76.5	9.9	211	5 PCT-US93-08364-16	Sequence 16, Appl

ALIGNMENTS

```

RESULT 1
US-08-233-788A-59
: Sequence 59, Application US/08233788A
: Patent No. 5635617
: GENERAL INFORMATION:
: APPLICANT: Doran, James L.
: APPLICANT: Kay, William W.
: APPLICANT: Collinson, Karen S.
: APPLICANT: Clouthier, Sharon C.
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
: TITLE OF INVENTION: OF SALMONELLA
: NUMBER OF SEQUENCES: 61
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Seed and Berry
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: Washington
: COUNTRY: U.S.A.
: ZIP: 98104-7092
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/233,788A
: FILING DATE: 26-APR-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: King, Joshua
: REGISTRATION NUMBER: 35,570
: REFERENCE/DOCKET NUMBER: 920043.403C2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: TELEX: 3723836 SEDANBERRY
: INFORMATION FOR SEQ ID NO: 59:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 151 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-233-788A-59

Query Match          99.4% Score 768: DB 1: Length 151:
Best Local Similarity 99.3% Pred. No. 8.8e-68:
Matches 150: Conservative 0: Mismatches 1: Indels 0: Caps 0:
1 MKLKVAFAAIVSGSALAGVPQMGCGGNHNGSGNSGPDSTLSIYOGSANAALALQ 60
|||||

```

Db 1 MLKLVAFATVYSGSALAGVVPQWGGGNNHNGGSSGPDSTLSIYOGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGAGADNSTIELTQNGFRNNATIDQWNAKSDITVGYCG 120
Db 61 SDARKSETTITQSGYNGADVGGAGADNSTIELTQNGFRNNATIDQWNAKSDITVGYCG 120
QY 121 NNALVNOTASDSSVMRYOYGFNNATANOY 151
Db 121 NNALVNOTASDSSVMRYOYGFNNATANOY 151

RESULT 2

US-08-233-788A-57
Sequence 57, Application US/08233788A
Patent No. 5635617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
TITLE OF INVENTION: OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible:
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANBERRY
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 120 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-233-788A-57

Query Match 76.1%; Score 588; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 2.3e-50;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 VVPQWGGGNNHNGGSSGPDSTLSIYOGSANAALALQSDARKSETTITQSGYNGADV 81
Db 1 VVPQWGGGNNHNGGSSGPDSTLSIYOGSANAALALQSDARKSETTITQSGYNGADV 60
QY 82 GCGADNSTIELTQNGFRNNATIDQWNAKSDITVGYOYGGNNALVNOTASDS 133
Db 61 GCGADNSTIELTQNGFRNNATIDQWNAKSDITVGYOYGGNNALVNOTASDS 112

RESULT 3
US-08-864-038A-3
Sequence 3, Application US/08864038A
Patent No. 6001592
GENERAL INFORMATION:

APPLICANT: Kunio NAKASHIMA et al.
TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: 812-5 Hirano
STREET: Isshinden
CITY: Tsu-city
STATE: Mie-prefecture
COUNTRY: JAPAN
ZIP: 514-01
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Word Perfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/864,038A
FILING DATE: May 28, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-184459
FILING DATE: 15-July-1996
ATTORNEY/AGENT INFORMATION:
NAME: C. Bruce Hamburg
REGISTRATION NUMBER: 22,389
REFERENCE/DOCKET NUMBER: F-5610
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)986-2340
TELEFAX: (212)953-7733
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 738
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Pinctada fucata
CELL TYPE: mantle epithelial cell
FEATURE:
NAME/KEY: peptide
LOCATION: from 1 to 738
IDENTIFICATION METHOD: E (by experiment)
US-08-864-038A-3

Query Match 12.6%; Score 97.5; DB 3; Length 738;
Best Local Similarity 27.5%; Pred. No. 0.16; 56; Indels 49; Gaps 6;
Matches 44; Conservative 11; Mismatches 33; Indels 49; Gaps 6;

QY 3 LLKVAFAFAIYSGSALAGVVPQWGGGNNHNGGSSGPDSTLSIYOGSANAALALQSD 62
Db 419 LLKSSASASASASASASASG-----GGGGGNGGNGGG-----GGAGALA---- 460
QY 63 ARKSETTITQSGYNGADVGGAGADNSTIELTQ-----NGFRNNATIDQWNAKSDITVGY 117
Db 461 -----AALAAAGAGCGGCGGCGALAAALAAAGAGCGGCGGCGG-----GG 503

QY 118 YCGNNALVNOTASDSS-----VWVROYGFNNATA 148
Db 504 LGGGSAALAAAAAASGSGGGRALRRALRRKGRGGGSAHA 543

RESULT 4
US-09-336-447A-5
Sequence 5, Application US/09336447A
Patent No. 6310190
GENERAL INFORMATION:
APPLICANT: HANSEN, ERIC J.
APPLICANT: AEBI, CHRISTOPH
APPLICANT: COPE, LESLIE D.
APPLICANT: MACIVER, ISOBEL

APPLICANT: FISKE, MICHAEL J.
APPLICANT: FREDENBURG, ROSS A.
TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
FILE REFERENCE: AMCY:024
CURRENT APPLICATION NUMBER: US/09/336,447A
CURRENT FILING DATE: 1999-06-21
NUMBER OF SEQ ID NOS: 98
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO: 5
LENGTH: 892
TYPE: PRT
ORGANISM: Moraxella catarrhalis
US-09-336-447A-5

Query Match 12.2%; Score 94; DB 4; Length 892;
Best Local Similarity 28.5%; Pred. No. 0.46;
Matches 41; Conservative 14; Mismatches 53; Indels 36; Gaps 8;

QY 28 GCGNHN-----GCGNNGPDETLSTIYOGSANAALALQSDARKSETTITQSGY---75
DB 89 CGGVNEAKGNVSTVGGSSMTAKGEKSTIGGGDTN-----DANGTYSTIGGTYGRA 141
QY 76 -GNGADVQGGADNSTI--ELTONGFRNNATIDONNAKNSDITVQYG--GNNALV---126
DB 142 IGDSSTIGGTYNOATGKSTVAGGRNN-----QATGNSTVAGSGSTVNOATGNSTVAGGS 197

OY 127 -NOTASDSSVMRYGVFGNNATAN 149
DB 198 HNOATGEGSF---AAGVENKANAN 218

RESULT 5
US-09-196-387-8
Sequence 8, Application US/09196387
Patent No. 6277613
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauder & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/196,387
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/095,225
FILING DATE: June 10, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 673 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-09-196-387-8

Query Match 11.9%; Score 92; DB 4; Length 673;
Best Local Similarity 30.4%; Pred. No. 0.5;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

OY 6 VAAFAAI-VVSGSALAGVYPOMGGGNNHNGGSSGPDSTLSIYOGSANAALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAFNPAGSGSNSSPSSPTSS--SSSPSPSGSLAESPPAA 157

OY 65 KSETTIT---OSGYGADVQGGADNSTITLQNG--FRNNATIDONNAKNSDI 113
DB 158 GVSSTAPLPGAGPGTGVPAVSGALRELLACRNGDVSRYKRLVDANVANAKDM 212

RESULT 6
US-09-196-387-10
Sequence 10, Application US/09196387
Patent No. 6277613
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauder & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/196,387
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/095,225
FILING DATE: June 10, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 949 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-196-387-10

Query Match 11.9%; Score 92; DB 4; Length 949;
Best Local Similarity 30.4%; Pred. No. 0.77;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

OY 6 VAAFAAI-VVSGSALAGVYPOMGGGNNHNGGSSGPDSTLSIYOGSANAALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAFNPAGSGSNSSPSSPTSS--SSSPSPSGSLAESPPAA 157

OY 65 KSETTIT---OSGYGADVQGGADNSTITLQNG--FRNNATIDONNAKNSDI 113
DB 158 GVSSTAPLPGAGPGTGVPAVSGALRELLACRNGDVSRYKRLVDANVANAKDM 212

```

; RESULT 7
; US-09-196-387-2
; Sequence 2, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OR USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; US-09-196-387-2

Query Match      11.9%; Score 92; DB 4; Length 1327;
Best Local Similarity    30.4%; Pctd. No. 1.2;
Matches   35; Conservative   15; Mismatches   57; Indels    8; Gaps

QY     6 VAAFAAI-VVSGALGAVVPOMGGGNHNGGNSGPDSTLSIYOGSANAIALAQSDAR 64
       ||| | :||| :||| | :||| :||| | :||| | :||| | :||| | :||| | :|||
Db     99 VAAPAVPVAVTSSTAAGVAPNPAGSSSNSSPSSSSPTFS -SSSPSSPSGSSLAESEPA 157
           :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY     65 KSETTIT---DSGYNGADVCGADNSTIELTQNG--FRNNATIDOWNAKKNDI 113
           :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db     158 GVASTAFLPGGAAGPCTGVPAVVSGLARELLEACRDNVDVRVKRLVDANVANAKDM 212

RESULT 8
US-09-336-447A-13
Sequence 13, Application US/09336447A
Patent No. 6310190
GENERAL INFORMATION:
APPLICANT: HANSEN, ERIC J.
APPLICANT: AEBEL, CHRISTOPH
APPLICANT: COPE, LESLIE D.
APPLICANT: MACIVER, ISOBEL
APPLICANT: FISKE, MICHAEL J.
APPLICANT: FREDENBURG, ROSS A.
TITLE OF INVENTION: USP#1 AND USP#2 ANTIGENS OF MORAXELLA CATARRHALIS

```

```

: FILE REFERENCE: AMCY:024
: CURRENT APPLICATION NUMBER: US/09/336,447A
: CURRENT FILING DATE: 1999-06-21
: NUMBER OF SEQ ID NOS: 98
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 13
: LENGTH: 873
: TYPE: PR1
: ORGANISM: Moraxella catarrhalis
US-09-336-447A-13

Query Match                      11.6%; Score 89.5; DB 4; Length 873;
Best Local Similarity 26.5%; Pred. No. 1.2;
Matches 40; Conservative 20; Mismatches 54; Indels 37; Gaps 9;

QY 4 LKVAFAIVVSGSLAGVYPQWGGGHHN--GGNSGSPSTLSTIYQGSANALALS 61
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 34 LIIVGLIMATTASQQTARQ--GKGHSLITIGGNDNEANGDYSTYSGGYN----- 86
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 DAKSETTITGSGYNGADVGQAGADNSTIELTONGFRNNATIDOMAKNSDITVQYGCN 121
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 87 ---KDSSTIGGGYNNEN-----GDSSTI---GGGFY-----EAKGESSTIG--GGD 127
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 NAALVNGTASDSSVYVRQVFG--NNATANQY 151
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 128 N-----NSATGMYSTIGGGDNNSATGRY 150
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
US-09-056-556-204
: Sequence 204, Application US/09056556
: Patent No. 6350456
: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Dillon, Davin C.
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
: NUMBER OF SEQUENCES: 241
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED and BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: Washington
: COUNTRY: USA
: ZIP: 98104-7092
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/056,556
: FILING DATE: 07-APR-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Makl, David J.
: REGISTRATION NUMBER: 31,392
: REFERENCE/DOCKET NUMBER: 210121.457
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO: 204:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 943 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
US-09-056-556-204

Query Match                      11.5%; Score 89; DB 4; Length 943;
Best Local Similarity 25.3%; Pred. No. 1.5;
Matches 39; Conservative 11; Mismatches 54; Indels 50; Gaps 7;

```

TRE


```

REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1864 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-804-227C-3

Query Match          10.5%; Score 81; DB 2; Length 1864;
Best Local Similarity 25.8%; Pred. No. 22;
Matches 34; Conservative 13; Mismatches 47; Indels 38; Gaps 4;

QY 6 VAAFPAT-----VYSGALAGVPPQWGGGNNHNGGNSGPDSTLSIYQGSANAALALQ 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 435 VSFSGAGTNAHVLESPVAGEPPAGRPEDTGGAWTVSG-----RGPAALR 481

QY 61 SDARKSETTITQSGYGADYGAGADNSTIEL-----TONGFRNNATIDQNNKNSDIT 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 482 AQAALYDALTGTTGTGGAGGAGPETAAYAGALAHARTAFRHRAVY----- 530

QY 115 VGOYGNNAALV 126
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 531 ---LGNRAELL 539

RESULT 13
US-08-737-716-2
Sequence 2, Application US/08737716
Patent No. 5955258
GENERAL INFORMATION:
APPLICANT: Girde BUIST
APPLICANT: Gerard VENEMA
APPLICANT: Jan KOK
APPLICANT: Adrianus Marinus LEDEROER
TITLE OF INVENTION: Process for the lysis of a culture of lactic
TITLE OF INVENTION: acid bacteria by means of a lysis, and uses of the resulting
TITLE OF INVENTION: lysed culture.
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/737,716
FILING DATE: 22-APR-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/ML95/00170
FILING DATE: 12-MAY-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: EP 94201353.3
FILING DATE: 12-MAY-1994
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 437 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-737-716-2

Query Match          10.4%; Score 80.5; DB 2; Length 437;
Best Local Similarity 21.2%; Pred. No. 3.7;

```

```

Matches 41; Conservative 32; Mismatches 53; Indels 67; Gaps 10;

QY 16 GSALAGVPPQW-----GGGNNHNGG-----NSGPDSTLSIYQGSANA----- 55
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 198 GAEINRIISQYNLFRPDGASACNTNSGSGTITTTNNNSGTSSTTYVKSGLTLMGIS 257

QY 56 -----ALALQSDARKSETTITQSGY-----GNGADYGAGADNSTIELT----- 93
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 258 QRYGIVAQIOISANNLSTIYIGOKLVLTGSASSTNSGSSNNSASTTPTTSTPAKPPS 317

QY 94 -----ONG-----FNNN-ATIDQNNKNSD-ITVGOYGNNAALVNTQASDSSVM 136
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 318 QTYVKAIVSGDTLALALSKYKTSIAQLKSMNHLSSDTIYIQ-----NLIVQSAASN-- 370

QY 137 VROYFGNNATAN 149
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 371 -PSTGSGSTATTNN 382

RESULT 14
US-09-043-123-2
Sequence 2, Application US/09043123A
Patent No. 6096521
GENERAL INFORMATION:
APPLICANT: HAAS, Rainer
APPLICANT: ODENBREIT, Stefan
APPLICANT: MEYER, Thomas F.
APPLICANT: BLUM, Andre
APPLICANT: CORTHESEY-THEULAZ, Irene
TITLE OF INVENTION: NEW ADHESIN FROM HELICOBACTER PYLORI
FILE REFERENCE: 05648004
CURRENT APPLICATION NUMBER: US/09/043,123A
CURRENT FILING DATE: 1998-06-26
EARLIER APPLICATION NUMBER: DE/195 35 321.8
EARLIER FILING DATE: 1995-09-22
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 518
TYPE: PRT
ORGANISM: Helicobacter pylori
US-09-043-123-2

Query Match          10.4%; Score 80.5; DB 3; Length 518;
Best Local Similarity 21.9%; Pred. No. 4.7;
Matches 37; Conservative 25; Mismatches 52; Indels 55; Gaps 6;

QY 29 GNNHNGGNSGPDSTLSIYQGSANA-----ALALQSDARKS 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 234 GPSGDSGAALFLDALAQHFNSNAGNDLSAKFTSLVQNIYVNSQNALTLNANNIS 293

QY 67 ETTITQSGYGADYGAGADNSTIELTONGFRNNATIDQNNKNSDIT---VGOYGN 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 294 NSTGYQVYGGID--QARSTOLLNNTN-----TLAKVALNLELKNPWLGNFAAGN 345

QY 123 AALVNO-----TASDSSVMROYGF-----GNNATANO 150
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 346 SSOVNAFMGFTIKIGYKQFFGKNKVGRLYYGFYSYNGAGVNGEPTINO 394

RESULT 15
US-08-209-747-8
Sequence 8, Application US/08209747
Patent No. 5733771
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: COLGIN, Mark
TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider
TITLE OF INVENTION: Silk Proteins
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747

```


THIS PAGE BLANK (USPTO)

Db 121 NNAALVNOTASDSSVYRVQVFGNNATANQY 151

RESULT 2

AI0635

major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium

C:Species: Salmonella enterica subsp. enterica serovar Typhimurium

A:Note: this species has also been called Salmonella typhimurium

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001

C:Accession: AI0635

R:Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Main, J., Churcher, C., Connor, T., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N., Farrar, S., Mout, S., O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., et al.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A:Reference number: AB0502; PMID:11677608

A:Accession: AI0635

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:916502315; GSPDB:GN00176

C:Genetics:

A:Gene: STY1181

Query Match

Best Local Similarity 100.0%; Score 773; DB 2; Length 151;

Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MMLKVAFAAIVVSGSALAGVPPQMGCGNHNGGNSGPDSTLSTIYQGSANALALQ 60

1 MMLKVAFAAIVVSGSALAGVPPQMGCGNHNGGNSGPDSTLSTIYQGSANALALQ 60

Db 1 MMLKVAFAAIVVSGSALAGVPPQMGCGNHNGGNSGPDSTLSTIYQGSANALALQ 60

Qy 61 SPARKSETTITGSGYNGADVQGGADNSTIELTONGFRNNATIDQWNAKNSDITVGYGG 120

61 SPARKSETTITGSGYNGADVQGGADNSTIELTONGFRNNATIDQWNAKNSDITVGYGG 120

Db 61 SPARKSETTITGSGYNGADVQGGADNSTIELTONGFRNNATIDQWNAKNSDITVGYGG 120

Qy 121 NNAALVNOTASDSSVYRVQVFGNNATANQY 151

121 NNAALVNOTASDSSVYRVQVFGNNATANQY 151

Db 121 NNAALVNOTASDSSVYRVQVFGNNATANQY 151

RESULT 3

S70788

curlin protein csqA precursor - Escherichia coli

N:Alternate names: csqA protein; major curlin protein

C:Species: Escherichia coli

C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 21-Jul-2000

C:Accession: S70788; G64846; S31202; S34560; S34559

R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csq operons is required for production of fliPronectin- and CsgA

A:Reference number: S70783; MUID:96414468

A:Accession: S70788

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:91147558; PIDN:CAA62282.1; PID:91147564

A:Experimental source: strain K12, substrain W3110

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC

A:Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: G64846

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:U00096; NID:91787265; PIDN:AAC74126.1; PID:91787279;

R:Olsen, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A:Title: The Rpos sigma factor relieves H-NS-mediated transcriptional repression of csgA

A:Reference number: S31202; MUID:93211294

A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6, 'V', 8-151 <OLS>

A:Cross-references: EMBL:L04979

A:Accession: S34560

A:Molecule type: Protein

A:Residues: 21-42;44-50 <OLS>

R:Olsen, A.N.; Arngvist, A.M.

submitted to the EMBL Data Library, October 1992

A:Reference number: S34559

A:Accession: S34559

A:Molecule type: DNA

A:Residues: 1-133, 'RGRDSGWLW' <OLS>

A:Cross-references: EMBL:L04979; NID:9290424; PIDN:AAA23616.1; PID:9290425

A:Experimental source: strain K-12, substrain W3110

C:Genetics:

A:Gene: csqA

A:Map position: 23.15

A:Function:

A:Description: major component of wild-type curli; interaction between CsgA and CsgB

A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli th

F1-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: curlin #status experimental <MAT>

Query Match

Best Local Similarity 75.8%; Score 586; DB 2; Length 151;

Matches 113; Conservative 19; Mismatches 19; Indels 0; Gaps 0;

Qy 1 MMLKVAFAAIVVSGSALAGVPPQMGCGNHNGGNSGPDSTLSTIYQGSANALALQ 60

1 MMLKVAFAAIVVSGSALAGVPPQMGCGNHNGGNSGPDSTLSTIYQGSANALALQ 60

Db 1 MMLKVAFAAIVVSGSALAGVPPQMGCGNHNGGNSGPDSTLSTIYQGSANALALQ 60

Qy 61 SPARKSETTITGSGYNGADVQGGADNSTIELTONGFRNNATIDQWNAKNSDITVGYGG 120

61 SPARKSETTITGSGYNGADVQGGADNSTIELTONGFRNNATIDQWNAKNSDITVGYGG 120

Db 61 SPARKSETTITGSGYNGADVQGGADNSTIELTONGFRNNATIDQWNAKNSDITVGYGG 120

Qy 121 NNAALVNOTASDSSVYRVQVFGNNATANQY 151

121 NNAALVNOTASDSSVYRVQVFGNNATANQY 151

Db 121 NNAALVNOTASDSSVYRVQVFGNNATANQY 151

RESULT 4

D90806

curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RI

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: D90806

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C

gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: D90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA834843.1; PID:913360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: EcsA420

Query Match

Best Local Similarity 73.0%; Score 564.5; DB 2; Length 152;

Matches 111; Conservative 20; Mismatches 20; Indels 1; Gaps 1;

Qy 1 MMLKVAFAAIVVSGSALAGVPPQMGCGNHNGGNSGPDSTLSTIYQGSANALALQ 59

1 MMLKVAFAAIVVSGSALAGVPPQMGCGNHNGGNSGPDSTLSTIYQGSANALALQ 59

Db 1 MMLKVAFAAIVVSGSALAGVPPQMGCGNHNGGNSGPDSTLSTIYQGSANALALQ 60

Qy 60 SPARKSETTITGSGYNGADVQGGADNSTIELTONGFRNNATIDQWNAKNSDITVGYGG 119

Db 61 QADARNSDLTITTOHGCGGAGADVGCGSDSDSLDTQRFGNSTALTDQWNGKSDHMTVKQFG 120
 Oy 120 GNNALVNQATASDSSVMVROVGFNNATANOV 151
 Db 121 GGNGAAYVDOTASNSTVNTVOGFGNNATAHOY 152

RESULT 5

hypotheetical protein csga [imported] - Escherichia coli (strain O157:H7, substrain EDL93 H85665
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: H85665
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Llm, A.; Diallanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: H85665
 A:Molecule type: DNA
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-152 <STO>
 A:Cross-references: GB:AE005174; NID:q12514574; PIDN:AAG55788.1; GSPDB:GN00145; UWGP:Z16
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: csga

Query Match

Best Local Similarity 73.0%; Score 564.5; DB 2; Length 152;
 Matches 111; Conservative 20; Mismatches 20; Indels 1; Gaps 1;

Oy 1 MKLLVAAFAATVSGSALAGVVPQW-GGGGNHNGCGNSGPDSTLSIYOGSANAALAL 59
 Db 1 MKLLVAAFAATVSGSALAGVVPQYGGGGHNGCGNSGNSSELNTIYOGGSALAL.60
 Oy 60 OSDARKSETTITQSGYGAGADVGAGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOY 119
 Db 61 QADARNSDLTITTOHGCGGAGADVGCGSDSDSLDTQRFGNSTALTDQWNGKSDHMTVKQFG 120
 Oy 120 GNNALVNQATASDSSVMVROVGFNNATANOV 151
 Db 121 GGNGAAYVDOTASNSTVNTVOGFGNNATAHOY 152

RESULT 6

hypotheetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) magA
 C:Species: Sinorhizobium meliloti
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C:Accession: E95965
 R:Flanagan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernan
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo
 A:Reference number: A95842; MUID:21396508; PMID:11481431
 A:Accession: E95965

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2174 <KUR>
 A:Cross-references: GB:AL591985; PIDN:CAC49389.1; PID:q15140875; GSPDB:GN00167
 A:Experimental source: strain 1021, megaplasmid pSymb
 R:Galibert, F.; Flanagan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
 pelat, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, J.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: Smb21548
 A:Genome: plasmid

Query Match 14.6%; Score 113; DB 2; Length 2174;
 Best Local Similarity 30.7%; Pred. No. 0.4;
 Matches 40; Conservative 21; Mismatches 51; Indels 36; Gaps 7;

Oy 11 AIVVSGSALAGVVPQ-WGGGNHNGCGNSGPDSTLSIYOGSANAALALQSDARKSET 68
 Db 693 AIVVSGSALAGVVPQWGGGNHNGCGNSGPDSTLSIYOGSANAALALQSDARKSET 68
 Oy 69 TTTQSGYGGAGDVG-----QGADNSTI--ELTQNGFRNNATIDQWNAKNSDITV 115
 Db 738 ---GCTATFANVGFKGLTLTTQSGSHAGIVQSVGGGGTGTASSYSAGIGFTAVAV 793
 Oy 116 GOYGNNAALVNQATASDSSVMVROVGFNNATANOV 141
 Db 794 GGTGNGAGAGVSVSLDSDAIRTGQGG 821

RESULT 7

S70787

curlin nucleator protein csqb precursor - Escherichia coli

N:Alternate names: csqb protein, curlin nucleation component; minor curlin protein

C:Species: Escherichia coli

C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 21-Jul-2000

C:Accession: S70787; F64846

R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csq operons is required for production of fibronectin- and

A:Reference number: S70783; MUID:96414468

A:Accession: S70787

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:q1147558; PIDN:CAA62281.1; PID:q1147553

A:Experimental source: Strain K12, substrain W3110

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: F64846

A:Molecule type: nucleic acid sequence not shown; translation not shown

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:q1787265; PIDN:AAC74125.1; PID:q17872

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: csqb

A:Map position: 23.15

C:Function:

A:Description: minor component of wild-type curli; interaction between CsgA and CsgB

A>Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli th

and H-Kinogen; in the absence of CsgA, CsgB can self-assemble into polymers

F:1-151/Domain: signal sequence #status predicted <STG>

F:12-151/Product: minor curlin chain #status predicted <MNT>

Query Match 14.6%; Score 112.5; DB 2; Length 151;

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 38 SSGPSTSTIYOGSANAALALQSDARKSETTITQSGYGAGADVGAGADNSTIELTQNGF 97
 Db 21 AAGYDLANSEYF---AVNELSKSSFNQAAITIGAGTNNKSQLQGGSKLLAAVVAQBS 76

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

RESULT 8

Query Match 14.6%; Score 112.5; DB 2; Length 151;

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASANDASISQAGAYGNFTAITIKSGGNKANTITQY 129

Query Match

Best Local Similarity 30.7%; Pred. No. 0.023;

Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

Oy 98 RNNATIDQWNAKNSDITVGOYGNNAALVNQATASDSSVMVROVGFNNATANOV 151
 Db 77 SRAKIDQGVNLAAYIDQASAND

```
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: C90806
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
  Sasawara, N.; Yasunaga, T.; Kuhara, S.; Shibba, T.; Hatori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: C90806
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA834842.1; PID:g13360879; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECs1419

Query Match          14.6%; Score 112.5; DB 2; Length 151;
Best Local Similarity 30.7%; Pred. No. 0.023;
Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

OY 38 SSGPSTLSTIYGSANALALQSDARKSETTITQSGYGNQADVGADNSTIELTQNGF 97
DB 21 AAGYDLANSEYF---AVNELSKSSEFNOAALIGQAGTNNQAOLKQSGSKLLAVVAQEGS 76
OY 98 RNNATIDQNNAKNSDITVGOYGGNNAALVNOTASDSSVWVRQVFGNNATANQY 151
DB 77 SRRAKIDQTDGVDNL-AYIDQAGSANDASISQAGYNTAMIIQKSGSKNANITQY 129

RESULT 9
curlin minor chain precursor, CsgA homolog [Imported] - Escherichia coli (strain O157:H7
G85665
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: G85665
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
  Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
  Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: G85665
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <STO>
A:Cross-references: GB:AE005174; NID:g12514573; PIDN:AAG55787.1; GSPDB:GN00145; UWGP:Z16
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: csgB

Query Match          14.6%; Score 112.5; DB 2; Length 151;
Best Local Similarity 30.7%; Pred. No. 0.023;
Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

OY 38 SSGPSTLSTIYGSANALALQSDARKSETTITQSGYGNQADVGADNSTIELTQNGF 97
DB 21 AAGYDLANSEYF---AVNELSKSSEFNOAALIGQAGTNNQAOLKQSGSKLLAVVAQEGS 76
OY 98 RNNATIDQNNAKNSDITVGOYGGNNAALVNOTASDSSVWVRQVFGNNATANQY 151
DB 77 SRRAKIDQTDGVDNL-AYIDQAGSANDASISQAGYNTAMIIQKSGSKNANITQY 129

RESULT 10
fimbria protein agfB precursor - Salmonella enteritidis
C:Species: Salmonella enteritidis
C:Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
C:Accession: J06040
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
  J. Bacteriol. 178, 662-667, 1996
A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.
A:Reference number: J06039; MUID:96146512
```

```
A:Accession: J06040
A:Molecule type: DNA
A:Residues: 1-151 <COL>
A:Cross-references: GB:U43280; NID:g1184712; PIDN:AAC43598.1; PID:g1184713
A:Experimental source: strain 276755-3b
C:Genetics:
A:Gene: agfB
A:Function:
A:Description: minor component of thin aggregative fimbriae
A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C:Keywords: fimbria
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-151/Product: fimbria protein agfB #status predicted <MAT>

Query Match          14.4%; Score 111.5; DB 2; Length 151;
Best Local Similarity 30.5%; Pred. No. 0.027;
Matches 36; Conservative 17; Mismatches 46; Indels 19; Gaps 5;

OY 51 GSANALALQSDARKSE---TTITQSGYGNQADVGQ-GADNST-----IELTQ 94
DB 14 GAPGIATATNYDLARSEYFAVNELSKSSEFNOAALIGQVGTNDRARVROEGSKLLSVISQ 73
OY 95 NGFRNNATIDQNNAKNSDIT-VGOYGGNNAALVNOTASDSSVWVRQVFGNNATANQY 151
DB 74 EGNRRARVQD--AGNYFAVIEQTGNANDASISQSAVNSAALIQKSGSKNANITQY 129

RESULT 11
nucleation component of curlin monomers [Imported] - Salmonella enterica subsp. enter
AH0635
C:Species: Salmonella enterica subsp. enterica serovar Typh
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C:Accession: AH0635
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
  th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
  S.; Moutle, S.; O'Goara, P.
  Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
  A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
  A:Reference number: AB0502; PMID:11677608
A:Accession: AH0635
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:g16502314; GSPDB:GN00176
C:Genetics:
A:Gene: STY1180

Query Match          14.4%; Score 111.5; DB 2; Length 151;
Best Local Similarity 30.5%; Pred. No. 0.027;
Matches 36; Conservative 17; Mismatches 46; Indels 19; Gaps 5;

OY 51 GSANALALQSDARKSE---TTITQSGYGNQADVGQ-GADNST-----IELTQ 94
DB 14 GAPGIATATNYDLARSEYFAVNELSKSSEFNOAALIGQVGTNDRARVROEGSKLLSVISQ 73
OY 95 NGFRNNATIDQNNAKNSDIT-VGOYGGNNAALVNOTASDSSVWVRQVFGNNATANQY 151
DB 74 EGNRRARVQD--AGNYFAVIEQTGNANDASISQSAVNSAALIQKSGSKNANITQY 129

RESULT 12
cnjB protein - Tetrahymena thermophila
S42136
C:Species: Tetrahymena thermophila
C:Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 07-Dec-1999
C:Accession: S42136; S42135; S03650
R:Taylor, F.M.; Martindale, D.W.
  submitted to the EMBL Data Library, October 1992
A:Reference number: S42136
A:Accession: S42136
A:Molecule type: DNA
```


A:Residues: 1-1748 <TAY>
A:Cross-references: EMBL:L03710; NID:g161751; PID:g161752
R:Taylor, F.M.; Martindale, D.W.
Nucleic Acids Res. 21, 4610-4614, 1993
A:Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by c
A:Reference number: 542135; MUID:9405156
A:Accession: 542135
A:Molecule type: DNA
A:Residues: 1164-1174;1179-1198;1233-1252;1285-1293;1297-1309;1316-1326;1331-1343-1
A:Cross-references: EMBL:L03710
R:Martindale, D.W.; Taylor, F.M.
Nucleic Acids Res. 16, 2189-2201, 1988
A:Title: Multiple introns in a conjugation-specific gene from *Tetrahymena thermophila*.
A:Reference number: S03650; MUID:88189811
A:Accession: S03650
A:Molecule type: DNA
A:Residues: 236-250, 'T', 252-255, 'N', 257-773 <MAR>
A:Cross-references: EMBL:X06462
C:Genetics:
A:Gene: cnjB
A:Genetic code: SGC5
A:Introns: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3; 8
C:Keywords: zinc finger
F:1164-1450/Region: glycine-rich
F:1451-1464/Region: zinc finger CCHC motif
F:1478-1491/Region: zinc finger CCHC motif
F:1501-1514/Region: zinc finger CCHC motif
F:1530-1543/Region: zinc finger CCHC motif
F:1555-1568/Region: zinc finger CCHC motif
F:1579-1592/Region: zinc finger CCHC motif
F:1602-1615/Region: zinc finger CCHC motif
F:1626-1748/Region: glycine-rich

Query Match 14.0%; Score 108; DB 2; Length 1748;
Best Local Similarity 30.5%; Pred. No. 0.8;
Matches 39; Conservative 20; Mismatches 29; Indels 40; Gaps 8;
QY 25 OMGGGNNHNGC--GNSGPDSTLSIYQGSANALALQSDARKSETTTTQSGYGADY*81
DB 1640 OFGGGNSNGGSGWMTSSSDNN-----CQSNVQES-TTSSSGGWS----- 1680
QY 82 GCGADNSTELTONGFRNNATIDQNNAKNSDITVGOYGGNNALVNOTASDSSVMYROYG 141
DB 1681 -SGSNQ---TGGGWSN---DNOQOQMENTGGGWSNS---NOTWNESS----- 1722
QY 142 FGNNTATN 149
DB 1723 WGSNNQAS 1730

RESULT 13
F70825
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 21-Jan-2000
C:Accession: F70825
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: F70825
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-645 <COL>
A:Cross-references: GB:AL021958; GB:AL123456; NID:g3261536; PID:CAAI7522.1; PID:e125329
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE
C:Superfamily: unassigned collagens

Query Match 13.9%; Score 107.5; DB 2; Length 645;
Best Local Similarity 25.8%; Pred. No. 0.29;
Matches 41; Conservative 17; Mismatches 66; Indels 35; Gaps 6;
QY 15 SSSALAGVPO-----WGGGNNHNGGNSGPDSTLSIYQGSANALALQSDARKSETT 69
DB 237 SCGNVIGIPSSFFNWSGNGNANWGCNSG--DNNFGFNFNANIGICNAGPNMSSPAV 295
QY 70 IQQS-----GYGADVQCGADNSTELTONGFRNNATIDQNNAKNSDITVGOY----- 119
DB 296 PTPGNGNWIGTGNNGNNGFGGNT-----GNANIGLVNGGNGVGFNSGSYNEGF 345
QY 120 ---GNNALVNOTASDS-----SVWYROYGFGNNATN 149
DB 346 GNTGNNNIGITGTSNQIFGGLNSGSGNIGFGNSTGN 364

RESULT 14
H98144
hypothetical protein AGR_L_228 [imported] - Agrobacterium tumefaciens (strain C58, Ce
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: H98144
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: H98144
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK8662.1; PID:g15158413; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_228
A:Map position: linear chromosome

Query Match 13.6%; Score 105.5; DB 2; Length 145;
Best Local Similarity 23.9%; Pred. No. 0.08;
Matches 37; Conservative 25; Mismatches 62; Indels 31; Gaps 3;
QY 3 LKVAFAFAIVVSGSALAGVPOWGC-----GNNHNGGNSGPDSTLSIYQGSANA 55
DB 1 MRRKSFIALVALVGLSAAPAMANDVRIEYGNSNSAGAOEGYGNRIITYONGCYNR 60
QY 56 ALALQSDARKSETTTTQSGYCGADVQCGADNSTELTONGFRNNATIDQNNAKNSDITV 115
DB 61 IVG-----HOYGRHMLSAVGQEGHDNYGSTTQNGNRNVAIGI----- 96
QY 116 GOYGNMALVNOTASDSSVMYROYGFGNNATN 150
DB 97 GQFGSNHTTTLTQDNGNIAAGVYGRGCSANVSQ 131

RESULT 15
AD3143
conserved hypothetical protein Atu4768 [imported] - Agrobacterium tumefaciens (strain
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AD3143
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo
kerage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; MCCI
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AD3143
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <KUR>
A:Cross-references: GB:AE008689; PIDN:AAL45562.1; PID:g17743277; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu4768
A:Map position: linear chromosome

Query Match 13.6%; Score 105.5; DB 2; Length 145;
Best Local Similarity 23.9%; Pred. No. 0.08;
Matches 37; Conservative 25; Mismatches 62; Indels 31; Gaps 3;

```
OY 3 LLKVAFAIIVSSSALAGVVPQWGG-----GGNHGCGNSGPDSTLSIYQGSANA 55
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MIRKSFIASALVALGLSAAAPAMANDVRIEYQWWSNSAGAGQEGYGNRIPTYONGYNR 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 56 ALALQDARKSETTITQSGYGNADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITY 115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 IVG-----HGYGRHLSAVGQEGCHDNVGTSTQNGNRNVAGT----- 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 116 GOYGNNAALVNOTASDSSVMVRQYVFGNNATANO 150
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 97 GQFGSNHTTILTQDNGNINAGVQVGRGCSANVSQ 131
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: October 11, 2002, 21:14:54
Job time : 22 secs

QY 86 DNSTELTONGFRNNATIDQWAKNSDITVQYGGNNALVNOTASDSSVMROYGFGNN 145
 Db 402 GNA-----GFFN---MGIGNAGNEDMGVNGSGSFNVGVGN--AGNOS-----VGFSGNA 444
 QY 146 ATAN 149
 Db 445 GTLN 448

RESULT 12
 P71868 PRELIMINARY; PRT; 582 AA.
 AC P71868.
 DT 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL. 55.5 KDA PROTEIN.
 GN RV3533C OR MTCY03C7.23.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Hornsby T., Jagels K., Patelwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers K.,
 RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RT Nature 393:537-544 (1998).
 RL EMBL: 282098; CAB05045.1;
 DR Tuberculest; RV3533C;
 DR InterPro: IPR002989; Mycobact.pentapep.
 DR InterPro: IPR000030; PPE.
 DR Pfam: PF01469; Pentapeptide_2; 10.
 DR Pfam: PF00823; PPE; 1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 582 AA; 55503 MW; B895BDD55979A9C6 CRC64;

Query Match 13.5%; Score 104.5; DB 16; Length 582;
 Best Local Similarity 28.6%; Pred. No. 4.6;
 Matches 40; Conservative 14; Mismatches 71; Indels 15; Gaps 6;

QY 23 VPQWGG--GNHGGGSSGPDSTLSIYQGSANALALOSDARKSETTITOSGNGAD 80
 Db 175 LPLMIGIGNIGNANLGGGNTG-DLNTGNGNIGNTNLGSGNNGDNLGSGNNGSNGV-GGN 232
 QY 81 VQD---GADSTEL-----TONGFRNNATIDQWAKNSDITVQYGGNNALVNOTASD 132
 Db 233 VGNNGNGSGNGRAGLPGSGNVGNGNLGNSLGSNGTNGNSVNGVGNNGNNVGTGNGNGSGN 292
 QY 133 ---SSVAVRQVGFNNATAN 149
 Db 293 IGAGNTGSSNWGFGNNGIGN 312

RESULT 13
 O9KRB1 PRELIMINARY; PRT; 1618 AA.
 AC O9KRB1;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE OMPB (FRAGMENT).
 GN OMPB.

OS Rickettsia japonica.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=35790;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YM;
 RX MEDLINE=20393643; PubMed=10939649;
 RA Roux V., Raoult D.;
 RT "Phylogenetic analysis of members of the genus Rickettsia using the
 RT gene coding the outer-membrane protein ompB (ompB)."
 RL Int. J. Syst. Evol. Microbiol. 50:1449-1455 (2000).
 DR EMBL: AF123713; AAF34116.1;
 DR InterPro: IPR003858; rompa.rompB.
 DR Pfam: PF02708; rompa.rompB; 1.
 FT NON_TER 1618
 SQ SEQUENCE 1618 AA; 164193 MW; 66AD2199620750C6 CRC64;

Query Match 13.3%; Score 103; DB 2; Length 1618;
 Best Local Similarity 26.2%; Pred. No. 18;
 Matches 45; Conservative 19; Mismatches 52; Indels 56; Gaps 8;

QY 6 VAAFAIVVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANALALOSDARK 65
 Db 477 VLAAGATLIDGSATL-----TGDTGNGG-----GAALSTITLANDATK 515
 QY 66 SEFTTQSG---YNGADVGOGADNSTELTONGFRNNATID----- 104
 Db 516 ---TLTGGANITSANGGTINFQANGGTLKLTST--QNNIYVDCCLATATDOTGVVDASS 570
 QY 105 QWAKNSDITVQYGGNNAL-----VNOTASDSSVMROYGFGNNAT 147
 Db 571 LNNAGTLTISGTIGTIGANNNTLGGFNGISKRTTLNGVNAINELVINGNS 622

RESULT 14
 O9SAF2 PRELIMINARY; PRT; 573 AA.
 AC O9SAF2;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE F3F19.21 PROTEIN.
 GN F3F19.21.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OX euroids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lee J.M., Lenz C.,
 RA Liu S., Li J., Kremetskaia I., Luros J., Ngan I., Gonzalez A.,
 RA Altieri H., Araujo R., Chao Q., Conn L., Conway A.B., Dunn P.,
 RA Hansen R., Huizlar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
 RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
 RT "Arabidopsis thaliana chromosome 1 BAC F3F19 sequence."
 RL Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Theologis A.;
 RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC007357; AAD31072.1;
 DR InterPro: IPR000504; RRM.
 DR Pfam: PF00076; Rrm; 1.
 DR SMART: SM00360; RRM; 1.
 DR PROSITE: PS50102; RRM; 1.
 SQ SEQUENCE 573 AA; 61400 MW; 0D1C10384B86C2BC CRC64;

Query Match 13.2%; Score 102; DB 10; Length 573;
 Best Local Similarity 26.1%; Pred. No. 6.9;

	OC	Bacteria;Proteobacteria; gamma subdivision; Enterobacteriaceae;
	OC	Escherichia.
	OX	NCBI_TaxId=562;
RN	[[]]	
RP	SEQUENCE FROM N.A.	
RC	TRANSPOSON-INSERTION SEQUENCE IS1;	
RX	MEDLINE=99314153; PubMed=10386375;	
RA	La Razione R.M., Colliphan R.J., Woodward M.J.:	
RT	"Non-curulation of Escherichia coli O78:K80 isolates associated with	
RT	IS1 inserti on in cs8B and reduced persistence in poultry infection."	
RL	FEMS Microbiol. Lett. 175:247-253(1999).	
DR	EMBL; AJ131756; CAB45377.1; -.	
FT	NON_TER	1
FT	NON_TER	91
SQ	SEQUENCE	91 AA; 9447 MW; E0C6FB91308CBDD CRC64;
	Query Match	13.6%; Score 105.5; DB 2; Length 91;
	Best Local Similarity	35.0%; Pred. No. 0.5;
	Matches	28; Conservative 9; Mismatches 42; Indels 1; Gaps
Dy	72 GSGGNGADVAGQGDNSTRIELTONGFRNNATIDQNNAKNSDITVGQYGNNALVNQTAS 131	
Dd	1 QAGTNNSAQQLQGGSKKLLAVVAOESSNRKATIDYGDYDNLAYIDQASANDASTSQGY 59	
OY	132 DSSVVRGVGFGNNTATANOY 151	
Dd	60 GNTAMTIIOKGSGKNRANITQY 79	
RESULT 11		
ID	053309 PRELIMINARY; PRT; 590 AA.	
AC	053309;	
DT	01-JUN-1998 (TREMBLrel. 06, Created)	
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)	
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE	PPE_FAMILY PROTEIN.	
GN	RVJ159C OR MTV014.03C.	
OS	Mycobacterium tuberculosis.	
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.	
OX	NCBI_TaxId=1773;	
RN	[[]]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=H37RV;	
RX	MEDLINE=98295987; PubMed=9634230;	
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,	
RA	Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,	
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,	
RA	Davies R., Devlin K., Dewell T., Gentles S., Hamlin N., Holroyd S.,	
RA	Hornsbly T., Jagielski K., Krogh A., McLean J., Moule S., Murphy L.,	
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,	
RA	Rutler S., Seeger K., Skelton S., Squares S., Squares R.,	
RA	Sustun J.E., Taylor K., Whitehead S., Barrell B.G.;	
RT	"Deciphering the biology of Mycobacterium tuberculosis from the	
RT	complete genome sequence."	
RL	Nature 393:537-544(1998).	
DR	EMBL; AL021646; CAI16624.1; -.	
DR	Tuberculist; RVJ159C; -.	
DR	InterPro; IPR002989; Mycobact_pentapep.	
DR	InterPro; IPR000030; PPE.	
DR	Pfam; PF01469; Pentapeptide_2; 10.	
DR	Pfam; PF00823; PPE; 1.	
KW	Complete proteome.	
SQ	SEQUENCE	590 AA; 56585 MW; 9AFC7F02FEABD588 CRC64;
	Query Match	13.6%; Score 105; DB 16; Length 590;
	Best Local Similarity	29.8%; Pred. No. 4.3;
	Matches	37; Conservative 15; Mismatches 46; Indels 26; Gaps
OY	26 WGGGHNHGCGNSSGPDSLTSLITYQSANAALALQSDARKSETTTIGSGGNGADVQGA 85	
Dd	351 FENSGNNNTGFPNSG--NNNVGFNSGNNNFEGCNACD-----INTGFGNAGDTNTTG 401	

Best Local Similarity 27.0%; Pred. No. 4.7;
Matches 40; Conservative 21; Mismatches 51; Indels 36; Gaps 7;

OY 11 ALVYSGSLAGVPP--WGGGNNHGGNSSGPDSTLSIYQGSANMALALOSDARKSET 68
 DB 693 AIAATAGAGAVIILAQISIGGGNN---GGNATGAGAGFSGFQIGGGGG-----737
 OY 69 TITGSGYNGADV-----OGADNSTI--ELTQNGFRNNATIDQWNAK---NSDITV 115
 DB 738 -----GGYANTANVGFKGLTTTQGSAAIYAQSVGGGGGTGCTASSISAGIGFTASVAV 793
 OY 116 GQYGNNA--LVNOTASDSSVMVROV 141
 DB 794 GGTGNGGAGGVSYSLTDSAIRTGOGG 821

RESULT 6

O33801 PRELIMINARY; PRT; 179 AA.
 AC O33801;
 DT 01-JAN-1998 (TREMblrel. 05, Last Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE AGFB PROTEIN.
 GN AGFB.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.
 NCBI_TaxID=602;
 RX MEDLINE=98053981; PubMed=9393832;
 RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeiffer J.D.,
 Normark S.J., Rhen M.;
 "Expression of chln, aggregative fimbriae promotes interaction of
 RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial
 cells.";
 RL Infect. Immun. 65:5320-5325(1997);
 DR EMBL: AJ000514; CA04150.1;
 SO SEQUENCE 179 AA; 19318 MW; A2BCB648B3C0B0B CRC64;

Query Match 14.1%; Score 109; DB 2; Length 179;
 Best Local Similarity 27.2%; Pred. No. 0.58;
 Matches 37; Conservative 19; Mismatches 48; Indels 32; Gaps 4;

OY 16 GSALAGVPPQGGCGNHNHGGSSGPDSTLSIYQGSANMALALOSDARKSETTITGSGY 75
 DB 63 GSKLSTVISO-ERGGNNRAKYDQAGNPFAYIEQTGNAN-----DASISQSAV 109
 OY 76 NGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITV-----GQYGNNA 123
 DB 110 GNSAI-----SAATIKGSGKNANITQYCTKTAVVYQKSHMALQANTQYCTGKT 162
 OY 124 ALVNOTASDSSVMVRO 139
 DB 163 AVVYQKSHMAIRVQ 178

RESULT 7

O93397 PRELIMINARY; PRT; 348 AA.
 AC O93397;
 DT 01-NOV-1998 (TREMblrel. 08, Created)
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE HYOSOPHORIN.
 OS Cypripus carpio (Common carp).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 CC Cypriniformes; Cyprinidae; Cyprinus.
 NCBI_TaxID=7962;
 RX NCI
 RN
 SO SEQUENCE FROM N.A.

RC TISSUE=OVARY;
 RX MEDLINE=99140211; PubMed=10206660;
 RA Tsao C.C., Huang F.L., Chang Y.S.;
 RT "Purification, characterization, and molecular cloning of carp
 RT hyosophorin";
 RL Mol. Reprod. Dev. 52:288-296(1999).
 DR EMBL: AF077819; AAC27329.1;
 SO SEQUENCE 348 AA; 33164 MW; 4CEAE146AB760035 CRC64;

Query Match 14.0%; Score 108.5; DB 13; Length 348;
 Best Local Similarity 27.4%; Pred. No. 1.3;
 Matches 40; Conservative 19; Mismatches 72; Indels 15; Gaps 5;

OY 16 GSALAGVPPQGGCGNHNHGGSSGPDSTLSIYQGSANMALALOSDARKSETTI 70
 DB 42 GSGSNGITQTDDGSGSNATTTTRDDGSGSNGITTR---DSSGSGNGITTTQDDGSGSNATT 97
 OY 71 TQ-SGYNGA---DVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYQYGNNAAL 125
 DB 98 TQDDGSGSNATTTQDDGSGSNATTTQDDGSGSNATTTQDDGSGSNATTTQDDGSGSNATT 157
 OY 126 VNQTRSDS-VVAVQYQYGNNAATANO 150
 DB 158 TQDDGSGSNGITTTQDDGSGSNATTTQ 183

RESULT 8

O94821 PRELIMINARY; PRT; 1748 AA.
 AC O94821;
 ID O94821; P92146; P92145; P92143; P92142; P92141; O94820;
 DT 01-FEB-1997 (TREMblrel. 02, Created)
 DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE CNB PROTEIN.
 GN CNB.
 OS Tetrahymena thermophila.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
 CC Tetrahymena; Tetrahymena.
 NCBI_TaxID=5911;
 RX MEDLINE=88189811; PubMed=3357771;
 RA Martindale D.W., Taylor F.M.;
 RT "Multiple introns in a conjugation-specific gene from Tetrahymena
 RT thermophila";
 RL Nucleic Acids Res. 16:2189-2201(1988).

RM [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94051569; PubMed=8233798;
 RA Taylor F.M., Martindale D.W.;
 RT "Retroviral-type zinc fingers and glycine-rich repeats in a protein
 RT encoded by cnb, a Tetrahymena gene active during meiosis";
 RL Nucleic Acids Res. 21:4610-4614(1993).
 DR EMBL: X06462; CAB37323.1;
 DR EMBL: I03710; AAC37171.1;
 DR HSSP; P05888; IAA.
 DR InterPro; IPR001878; ZnF_CCHC.
 DR Pfam; PF000098; zf-CCHC; 7.
 DR PRINTS; PR00939; CZHCZNFINGER.
 DR SMART; SM00343; ZnF_C2HC; 7.
 RW Zinc-finger.
 FT CONFLICT 251 M -> I (IN REF. 1).
 FT CONFLICT 256 I -> N (IN REF. 1).
 SO SEQUENCE 1748 AA; 199624 MW; 0B03F210104008A3 CRC64;

Query Match 14.0%; Score 108; DB 5; Length 1748;
 Best Local Similarity 30.5%; Pred. No. 8.5;
 Matches 39; Conservative 20; Mismatches 29; Indels 40; Gaps 8;

OY 25 OMGGGNNHGG---GNSGPDSTLSIYQGSANMALALOSDARKSETTITGSGYNGADV 81
 DB 1640 OFGGGNGNSGSGWGTSSGSDMN-----COSNVQES-TTSSGSGWS-----1680

Db 121 NNALVNOTASDSSVWRYOVGFNNAPANOY 151

RESULT 2

Q93024 PRELIMINARY; PRT; 152 AA.
AC Q93024;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE CSGA PROTEIN.
GN CSGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=43895 RED VARIANT;
RX MEDLINE=21218556; PubMed=11319125;
RA Ulrich G.A., Keen J.E., Elder R.O.;
RT "Mutations in the csgD promoter Associated with Variations in Curli
Expression in Certain Strains of Escherichia coli O157:H7."
RL Appl. Environ. Microbiol. 67:2367-2370(2001).
DR EMBL: AF275733; AAK53212.1; -
SQ SEQUENCE 152 AA; 15099 MW; E82DD94DDE91243 CRC64;

Query Match 73.0%; Score 564.5; DB 2; Length 152;
Best Local Similarity 73.0%; Pred. No. 2e-34;
Matches 111; Conservative 20; Mismatches 20; Indels 1; Gaps 1;

OY 1 MLLKVAFAAIVVSGSALAGVPOW-GGGGNHGGSGSPDSTLSIYOGSANAALAL 59
Db 1 MLLKVAFAAIVVSGSALAGVPOYGGGGGSGGNNSPENLTYOYGGNSALAL 60
OY 60 QSDARKSETTOSGCGADVGAGADNSTELTONGFRNNATIDQNNAKNSDITVGOYG 119
Db 61 QADARNSLTLTTHOGGAGADVGCGSDSDSIDLQRCFGNSATLTDQNNKSHMTVKQFG 120
OY 120 GNNALVNOTASDSSVWRYOVGFNNATANOY 151
Db 121 GGNGAAVDOTASNSTVNTYOVGFNNATANOY 152

RESULT 3

Q54069 PRELIMINARY; PRT; 76 AA.
AC Q54069;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE SEF17 FIMBRIN (FRAGMENT).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Eglezios S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
colony morphology and expression of SEF17 fimbriae."
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U53207; AAG98671.1; -
FT NON-TER 1
KW SEQUENCE 76 AA; 7704 MW; 2FDS411241A7BCD1 CRC64;
SQ

Query Match 49.8%; Score 385; DB 2; Length 76;
Best Local Similarity 98.7%; Pred. No. 1.3e-21;
Matches 75; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 30 GNHGGSSGSPDSTLSIYOGSANAALALQSDARKSETTITSGYCGNAGADVGADNST 89
Db 1 GNHGGSSGSPDSTLSIYOGSANAALALQSDARKSETTITSGYCGNAGADVGADNST 60

OY 90 IELTONGFRNNATIDQ 105
Db 61 IELTONGFRNNATIDQ 76

RESULT 4

Q9S3J5 PRELIMINARY; PRT; 29 AA.
AC Q9S3J5;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE CORLIN SUBUNIT MONOMER (FRAGMENT).
GN CSGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-INSERTION SEQUENCE ISI;
RX MEDLINE=99314153; PubMed=10386375;
RA La Ragione R.M., Collighan R.J., Woodward M.J.;
RT "Non-cultivation of Escherichia coli O78:K80 Isolates associated with
RT ISI inserti on in csgB and reduced persistence in poultry infection."
RL FEMS Microbiol. Lett. 175:247-253(1999).
DR EMBL: AJ131756; CAB45380.1; -
FT NON-TER 29
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABB8243 CRC64;

Query Match 15.8%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.0087;
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 MLLKVAFAAIVVSGSALAGVPOWGGG 29
Db 1 MLLKVAFAAIVVSGSALAGVPOYGGG 29

RESULT 5

Q92U08 PRELIMINARY; PRT; 2174 AA.
AC Q92U08;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HYPOTHETICAL GLYCINE-RICH PROTEIN SMB21548.
GN SMB21548.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymb (megaplasmid 2).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RA MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorholster F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puehler A.;
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
fixing endosymbiont Sinorhizobium meliloti."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR EMBL: AF603645; CAC49389.1; -
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 2174 AA; 203314 MW; 008BB68297B44182 CRC64;

Query Match 14.6%; Score 113; DB 16; Length 2174;

RX MEDLINE=87283911; PubMed=3302276;
 RA Riede I., Drexler K., Eschbach M.L., Henning U.;
 RT "DNA sequence of genes 38 encoding a receptor-recognizing protein of
 bacteriophages T2, K3 and of K3 host range mutants.";
 RL J. Mol. Biol. 194;31-39(1987).
 CC -1- FUNCTION: VG38 IS AT THE TIP OF THE LONG TAIL FIBERS AND SERVES
 CC AS THE PHAGE RECOGNITION SITE FOR THE CELLULAR RECEPTOR.
 CC -1- MISCELLANEOUS: THIS PHAGE USE OUTER MEMBRANE PROTEINS OMPF AND TTR
 CC AS RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X05312; CAA28935.1; -.
 DR PIR: S00275; S00275.
 KM Fiber protein; Phage recognition.
 SQ SEQUENCE 262 AA; 25801 MW; 0567366918F6C745 CRC64;

Query Match 11.6%; Score 89.5; DB 1; Length 262;
 Best Local Similarity 34.1%; Pred. No. 1.9;
 Matches 30; Conservative 8; Mismatches 33; Indels 17; Gaps 4;
 OY 27 GCGGNHNGCGSGPDSTLSIYQGSANALALQSDARKSETTITOSGYGNGADVGGGAD 86
 Db 175 GCGGRPGVCGKIGSDSILS-----GSNASL--TDAGTGCTF-QYGAGNGGNVAGAGG 225
 OY 87 NSTIELTONGFRNNTATIDWNANKNSDIT 114
 Db 226 -----RCGMRKNVYTSSEGAAGAAVT 245

Search completed: October 11, 2002, 21:13:42
 Job time : 13 secs

STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY SIMILARITY).

-I- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR (BY SIMILARITY).

-I- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-LAYER WITH HEXAGONAL SYMMETRY (BY SIMILARITY).

-I- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/or_send_an_email_to_license@sib-sib.ch).

DR EMBL; AE008659; AAL03623.1; -
DR EMBL; AF123721; AAF34124.1; -
DR EMBL; AF123726; AAF34128.1; -
DR EMBL; AF149110; AAD39533.1; -
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
KW Antigen; S-layer; Cell wall; Complete proteome.
FT CHAIN 1 1334
FT CHAIN 1 1334
FT CHAIN 1335 1655
FT VARIANT 61 61
FT VARIANT 75 75
FT VARIANT 78 78
FT VARIANT 251 251
FT VARIANT 413 413
FT VARIANT 959 959
FT VARIANT 988 988
FT VARIANT 1139 1139
FT CONFLICT 353 354
FT CONFLICT 776 776
FT CONFLICT 1159 1159
FT CONFLICT 1177 1177
FT CONFLICT 1492 1492
SO SEQUENCE 1655 AA; 168342 MW; E4E19377D5FCE37 CRC64;

Query Match 11.9%; Score 92; DB 1; Length 1655;
Best Local Similarity 25.6%; Pred. No. 8.6;
Matches 43; Conservative 17; Mismatches 68; Indels 40; Gaps 7;

QY 13 VVSSALAGVPPQWGGGNNHNGGSSPDBSLTIYQGSANALALD-----SDARKSE 67
Db 490 LADGVLNGKNGQNALVGCALAGTITLDSGATTTGDIAGNAGGAALOGITLANDATK-- 547
QY 68 TTITQSGY----GNGADVGAGADNSTIETLQNGFRNNATID-----QW 106
Db 548 -TLTLFGANITIGANGGTINFGANGGTIKLTST--QNNIVYDFDLAINTDQGVYDASSLT 604
QY 107 NAKMSDIT--VGQYGSNNAL-----VNOTASDSSVMVQVQFGNN 145
Db 605 NAQPLTITNGKIGTIVGANNKTLGQFNIGSSKTVLSDGVAINELVIGNN 652

RESULT 15
VG38_BPT2
ID VG38_BPT2 STANDARD; PRT; 262 AA.
AC P07875;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)
DE Receptor recognizing protein (Protein Gp38).
GN 38.
OS Bacteriophage T2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like phages.
OX NCBI_TaxID=10664;
RP [1]
SEQUENCE FROM N A

```

CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; M25382; AAA24823.1; -.
DR  PIR; J00188; J00188.
DR  HSSP; P06620; 1INA.
DR  InterPro; IPR000258; Ice_nucleatn.
DR  Pfam; PF00818; Ice_nucleation; 65.
DR  PRINTS; PR00327; ICENUCLEATN.
DR  PROSITE; PS00314; ICE_NUCLEATION; 45.
KW  Ice nucleation; Repeat; Outer membrane.
FT  DOMAIN 162 1217 OCTAPEPTIDE PERIODICITY.
SQ  SEQUENCE 1258 AA; 125084 MW; 590E8A130077EBD4 CRC64;

Query Match 12.4% Score 95.5; DB 1; Length 1258;
Best Local Similarity 27.5% Pred. No. 3.4; Mismatches 49; Indels 61; Gaps 14;
Matches 52; Conservative 27;

OY 6 VAAFAIVVSG---SALGV-VPOWGGGNH---NGGNSGPDSTLSIYOGSANAAL 57
DB 209 VAGVSTGTAGESSQMGVSTGTGMKGSDDLTAAGSTGTAGDSSSL- IAGVSTGTA- 266
OY 58 ALQSDARKESTTTTOSGYN-----GADV-----GGGADNSTI-----EL 92
DB 267 -----GEDSLTAGVSTGTAGKGSDDLTAAGVSTGTAGDSSSLIAGVSTGTAGEES 318
OY 93 TON-GERNNATIDQWNAKNSDITVGOYG-----GNNAALV-----NOTASDSSVWVROYG 141
DB 319 TOTAGVSTGTATQ-----KQSDLTAG-YGSTGTAGDSSSLIAGVSTGTAGESSSLT--AG 371
OY 142 FGNNTATANO 150
DB 372 YGSTGTATQK 380

RESULT 12
ID ICEA_PANAN STANDARD: PRT: 1322 AA.
AC P20469;
DT 01-FEB-1991 (Rel. 17, Created).
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein 1naa.
GN INAA.
OS Pantoea ananas (Erfwinia uredovora).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90092494; PubMed=259095;
RA Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;
RT "An ice nucleation active gene of Erfwinia ananas. Sequence similarity
RT to those of Pseudomonas species and regions required for ice
RT nucleation activity."
RT FEBS Lett. 258:297-300(1989).
RL
CC -1- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC CRYSTALLIZATION IN SUPERCOOLED WATER.
CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -1- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR  EMBL; X17316; CAA35194.1; -.
DR  PIR; S07053; S07053.
DR  HSSP; P06620; 1INA.
DR  InterPro; IPR000258; Ice_nucleatn.
DR  Pfam; PF00818; Ice_nucleation; 69.
DR  PRINTS; PR00327; ICENUCLEATN.
DR  PROSITE; PS00314; ICE_NUCLEATION; 49.
KW  Ice nucleation; Repeat; Outer membrane.
FT  DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.
SQ  SEQUENCE 1322 AA; 131094 MW; 89B0E24AA837039 CRC64;

Query Match 12.1% Score 93.5; DB 1; Length 1322;
Best Local Similarity 28.3% Pred. No. 5.2; Mismatches 24; Indels 49; Gaps 8;
Matches 36; Conservative 18;

OY 34 GGGNSGPDSTLSIYOGSANAALALQSDARKSETTTTOSGNGADVGGGADNSTIELT 93
DB 933 GSTTAPDSSL-IAGVSTGTA-----GYNSTLTAGYGS-----T 967
OY 94 QNGFRNATIDQWNAKNSDITVGOYG-----GNNAALV-----NOTASDSSVWVROYG 143
DB 968 QTG-----QENSDDLTTG-YGSTTAGYESSSLIAGVSTGTAFKSTLM--AGYG 1013
OY 144 NNATANO 150
DB 1014 SSGTARE 1020

RESULT 13
ID N145_YEAST STANDARD: PRT: 1317 AA.
AC P49687;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Nucleoporin NUP145 (Nuclear pore protein NUP145).
GN NUP145 OR YGL092W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94320139; PubMed=8044840;
RA Fabre E., Boelens W.C., Wimmer C., Matthej I.W., Hurt E.C.;
RT "Nup145p is required for nuclear export of mRNA and binds
RT homopolymeric RNA in vitro via a novel conserved motif."
RT Cell 78:275-289(1994).
RL
CC [2]
CC SEQUENCE FROM N.A.
CC STRAIN=ATCC 26109 / X2180;
CC MEDLINE=97435481; PubMed=9290212;
CC Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
CC "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
CC chromosome VII."
CC Yeast 13:1077-1090(1997).
CC -1- FUNCTION: MAY PLAY A DIRECT ROLE IN NUCLEOCYTOPLASMIC EXPORT OF
CC RNA AND ALSO IN PROTEIN IMPORT. BINDS HOMOPOLYMERIC RNA.
CC -1- SUBCELLULAR LOCATION: Nuclear pore complex.
CC -1- SIMILARITY: BELONGS TO THE GLEF FAMILY OF NUCLEOPORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; 215035; CAA78753.1; -
DR EMBL; X75780; CAA53406.1; -
DR EMBL; 228068; CAA81905.1; -
DR PIR; B44402; B44402.
DR PIR; S39173; S39173.
DR PIR; S4518; S4518.
DR SGD; S0001551; NUP100.
DR InterPro; IPR004325; Nucleoporin_Fg.
DR Pfam; PF03093; Nucleoporin_Fg; 24.
KW Nuclear protein; Transport; Repeat.
FT DOMAIN 33 571 29 X 6 AA APPROXIMATE REPEATS OF
FT G-L-F-G.
SQ SEQUENCE 959 AA; 99988 MW; D3985F9901BBA51 CRC64;

Query Match 12.4%; Score 96; DB 1; Length 959;
Best Local Similarity 27.4%; Pred. No. 2.3;
Matches 37; Conservative 17; Mismatches 53; Indels 28; Gaps 6;

QY 27 GGGGNHGGGNSGPDSTLSIV--QYGSANMALALQSDARKSETTITOSGYGN-----GA 79
DB 183 GNGSNIFGAGNNSQMTSTSLFNGQSSAFGTNNQGSLEFGQSSQNTNNAFGNQNLGSS 242
QY 80 DVGQ-----GADNSTIELTNGFRNNATIDQNNANSDITVQYCGNNALYNQTA 130
DB 243 SFQSKFVSGSLFGQSNNTLGNITT--NRNGLEFGQNMSSNQ-----GSSNSGLEFGQNS 293

QY 131 SDSSVAVRQVGFQNN 145
DB 294 MNSST---QGVTCQN 305

RESULT 10
ID ICEN_PANAN STANDARD; PRT: 1034 AA.
AC 047879;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein inau.
GN INAU.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=94264407; PubMed=7764866;
RA Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
RT "Cloning and sequencing of an ice nucleation active gene of Erwinia
RT uredovora.";
RL Biosci. Biotechnol. Biochem. 58:762-764(1994).
CC -I- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC CRYSTALLIZATION IN SUPERCOOLED WATER.
CC -I- SUBCELLULAR LOCATION: Outer membrane.
CC -I- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -I- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -I- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; D14992; BAA03636.1; -
DR HSSP; P06620; IINA.
DR InterPro; IPR000258; Ice-nucleatn.
DR Pfam; PF00818; Ice_nucleation; 51.
DR PRINTS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 34.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162 993 OCTAPEPTIDE PERIODICITY.
FT SEQUENCE 1034 AA; 103378 MW; FA222523D33EADD CRC64;

Query Match 12.4%; Score 95.5; DB 1; Length 1034;
Best Local Similarity 27.5%; Pred. No. 2.8;
Matches 52; Conservative 27; Mismatches 49; Indels 61; Gaps 14;

QY 6 VAAPFAIVVSG---SALGV-VPWGGGQNH-----NGGNSGSPDSTLSIYQGSANMAL 57
DB 209 VAGYSTGTAGESSQMGAYGSTGTGMKGSDLTAGYGTGTAGDSSL-IAGYSTGTA- 266
QY 58 ALQSDARKSETTITOSGYGN-----GADV-----GQGANSTI-----EL 92
DB 267 -----GEDSLTAGYGSTGTAGKGSDLTAGYGTGTAGADSSLLIAGYSTGTAGEES 318
QY 93 TGN-GEFNATIDQNNANSDITVQYQ-----GNMALV-----NQTSADSSVAVRQV 141
DB 319 TGTAGYGTGTATQ-----KGSDLTAG-YGSTGTAGDSSLIAGYGSTGTAGESSSLT--AG 371

QY 142 FGNNTANO 150
DB 372 YGSTGTAK 380

RESULT 11
ID ICEN_ERWHE STANDARD; PRT: 1258 AA.
AC P16239;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN ICEE.
OS Erwinia herbicola.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MJ.
RA MEDLINE=90152370; PubMed=2515997;
RA Warren G.J., Corotto L.V.;
RT "The consensus sequence of ice nucleation proteins from Erwinia
RT herbicola. Pseudomonas fluorescens and Pseudomonas syringae.";
RL Gene 85:239-242(1989).
CC -I- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC CRYSTALLIZATION IN SUPERCOOLED WATER.
CC -I- SUBCELLULAR LOCATION: Outer membrane.
CC -I- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -I- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -I- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

FT DOMAIN 797 802 POLY-GLN.
FT DOMAIN 820 823 POLY-GLN.
FT DOMAIN 826 832 POLY-GLN.
FT DOMAIN 874 992 ZINC FINGERS.
FT ZN_FING 874 896 C2H2-TYPE.
FT ZN_FING 902 924 C2H2-TYPE.
FT ZN_FING 930 933 C2H2-TYPE.
FT ZN_FING 969 992 A -> R (IN REF. 2).
FT CONFLICT 647 647
SQ SEQUENCE 1028 AA: 110620 MW: D7068BB2BC0F6F77 CRC64:

Query Match 12.7% Score 98.5; DB 1; Length 1028;
Best Local Similarity 28.7%; Pred. No. 1.6; Indels 35; Gaps 7;
Matches 43; Conservative 13; Mismatches 59;

OY 3 LKAAFAATVSSGSLAGVVPWGGGNNHGGNSGSPDSTLSTIYOGSAMAALALQSD 62
DB 59 LQNAALAAVYMSAGS-----GGGCTGNGGAGSAGPGGSPRANGSGGGG----- 104
OY 63 ARKSETTITSGYNGADVGADNGADNSTIELTQNGFRNATIDOMANKNSDI---TVGQY 118
DB 105 -----GGNGYINCAGVG-GRPNNS--LDGNMLNFASVSNYESNKRFFNNHHHHHH 152
OY 119 GGNNAALVNOTASDSVWVROYGFGNATA 148
DB 153 NNNN-----NNNGGQTSWGHGF-YGNGPSA 177

RESULT 8
YF48_MYCTU STANDARD: PRT; 678 AA.
ID YF48_MYCTU
AC Q10778;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PPE-family protein RV1548C.
GN RV1548C OR MT1599 OR MYC148.17.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium;
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson S.J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayan J.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bissel W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 274020; CAA9835.1;
DR EMBL: AE007026; AAK45866.1; ALT_INIT.
DR TIGR: MT1599;
DR TubercuList: RV1548C;
DR InterPro: IPR002989; Mycobact_Pentapep.
DR InterPro: IPR000030; PPE.
DR Pfam: PF01469; Pentapeptide_2; 13.
DR Pfam: PF00823; PPE; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 14 34
FT TRANSMEM 180 200 POTENTIAL.
FT CONFLICT 258 258 D -> G (IN REF. 2)
SQ SEQUENCE 678 AA: 66736 MW: 209F1593D52533A2 CRC64:

Query Match 12.7% Score 98; DB 1; Length 678;
Best Local Similarity 28.0%; Pred. No. 1.1;
Matches 37; Conservative 10; Mismatches 45; Indels 40; Gaps 8;

OY 26 WGGG--GNHN-GGNGSSGPDSTLSTIYOGSAMAALALQSDARKSETTITSGYNGADVG 82
DB 267 WGGGNGISYMLGGN-----LGSYMLGSGN-----TGDTFGGGNGTGNLN 306
OY 83 QGADNSTIELTON---GFRNATIDOMANKNSDITV---GGYGNNAALVNOTASDSVWV 137
DB 307 VGGGN-----TGNSNFFGFTGNTGVNGTGTGTFGSGNLGSGNIGFGNKG----- 354
OY 138 ROVGFGNATATN 149
DB 355 HNTFGNSGN 366

RESULT 9
N100_YEAST STANDARD: PRT; 959 AA.
ID N100_YEAST
AC Q02629;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Nucleoporin NUP100/NSP100 (Nuclear pore protein NUP100/NSP100).
GN NUP100 OR NSP100 OR YKL068W OR YKL336.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054906; PubMed=1385442;
RA Wente S.R., Rout M.P., Blobel G.;
RA "A new family of yeast nuclear pore complex proteins."
RL J. Cell Biol. 119:705-723(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=94378724; PubMed=8091863;
RA Rasmussen S.W.;
RA "Sequence of a 20.7 kb region of yeast chromosome XI includes the
RT NUP100 gene, an open reading frame (ORF) possibly representing a
RT nucleoside diphosphate kinase gene, tRNAs for His, Val and Tyr in
RT addition to seven ORFs with weak or no significant similarity to
RT known proteins."
RL Yeast 10:569-574(1994).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF
CC THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.
CC -1- SUBCELLULAR LOCATION: Nuclear pore complex.
CC -1- DOMAIN: CONTAINS G-L-F-G REPEATS.
CC -1- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```
RESULT 6
ICEN_XANCT STANDARD: PRT: 1567 AA.
AC P18127;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN INAX.
OS Xanthomonas campestris (pv. translucens).
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xanthomonas.
OX NCBI_TaxID=343;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-X565;
RX MEDLINE-91080859; PubMed-2259339;
RA Zhao J., Orser C.S.;
RT "Conserved repetition in the ice nucleation gene inax from
RT Xanthomonas campestris pv. translucens.";
RL Mol. Gen. Genet. 223:163-166(1990).
CC -1- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC CRYSTALLIZATION IN SUPERCOOLED WATER.
CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -1- DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-L-T: FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -1- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X52970; CA37140.1; -
DR PIR, S11672; S11672.
DR HSSP, P06620; 11NA.
DR InterPro: IPR000258; Ice_nucleatn.
DR Pfam: PF00818; Ice_nucleation; 81.
DR PRINTS: PR00327; ICENUCLEATN.
DR PROSITE: PS00314; ICE_NUCLEATION; 57.
DR Ice nucleation; Repeat; Outer membrane.
SQ SEQUENCE 1567 AA; 152548 MW; C8B451D959ECAD63 CRC64;

Query Match 13.1%; Score 101.5; DB 1; Length 1567;
Best Local Similarity 27.2%; Pred. No. 1.5;
Matches 49; Conservative 25; Mismatches 53; Indels 53; Gaps 12;

OY 14 VEGSALAG-----VVPQMG-----GGNMN-----GGGSSGPDSTSIYGSANMAL 59
DB 205 VVGSTLTGADQSRVAVGYSTETAGDHDLIAGYGTAGSDST-LAGYGTOTAGR 263
OY 60 Q-----SDARKSETTITQSGYGVGADVGAGADNSTIELTNGFRNNATI----- 103
DB 264 STLTAGYGTGTGAEGSRLLTSGYGTAT--SGSDSAVI-----SGYGTOTAGSESSLT 317
OY 104 ---DGMANKNSDITVGYQY-----GNMALV-----NOTASDSSWVAVRGGNATANO 150
DB 318 YGSTGTARKGSDITAG-YGSTGTAGSDSALLIAGYGTOTAGSESSLT--AGYGTOTARK 374

RESULT 7
OYO_DROME STANDARD: PRT: 1028 AA.
AC P51521; Q9XZU4;
DT 01-OCT-1996 (Rel. 34, Created)
```

```
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Oyo protein (shaven baby protein).
GN OYO OR SVB.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE-95021209; PubMed-7935398;
RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;
RT "Multiple products from the shavenbaby-ovo gene region of Drosophila
RT melanogaster: relationship to genetic complexity.";
RL Mol. Cell. Biol. 14:6809-6818(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE-91293102; PubMed-1712294;
RA Mevel-Minto M.T.M., Terracol R., Kafatos F.C.;
RT "The ovo gene of Drosophila encodes a zinc finger protein required
RT for female germ line development.";
RL EMBO J. 10:2259-2266(1991).
CC -1- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM
CC LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIIUM AND
CC ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG,
CC BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED
CC PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, 011383; AAB60216.1; -
DR EMBL, X59772; CAB36921.1; ALT_SEQ.
DR HSSP, P25490; 1ZNM.
DR FlyBase: FBgn003028; ovo.
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam: PF00096; Zf-C2H2; 4.
DR PRINTS: PR00048; ZINCFINGER.
DR SMART: SM00355; Znf_C2H2; 4.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 3.
DR Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KW Transcription regulation.
FT DOMAIN 62 77 POLY-ALA.
FT DOMAIN 72 77 POLY-GLY.
FT DOMAIN 80 85 POLY-GLY.
FT DOMAIN 98 108 POLY-GLY.
FT DOMAIN 144 152 POLY-HIS.
FT DOMAIN 153 159 POLY-ASN.
FT DOMAIN 336 339 POLY-GLN.
FT DOMAIN 347 353 POLY-GLN.
FT DOMAIN 357 361 POLY-GLN.
FT DOMAIN 410 414 POLY-GLN.
FT DOMAIN 418 422 POLY-GLN.
FT DOMAIN 426 432 POLY-GLN.
FT DOMAIN 445 453 POLY-GLN.
FT DOMAIN 456 459 POLY-GLN.
FT DOMAIN 466 474 POLY-GLN.
FT DOMAIN 497 517 POLY-ALA.
FT DOMAIN 524 529 POLY-SER.
FT DOMAIN 549 558 POLY-ALA.
FT DOMAIN 639 651 POLY-ALA.
FT DOMAIN 717 725 POLY-ALA.
```

```

01  RESULT 4
02  CSGB_SALT
03  ID CSGB_SALT STANDARD; PRT; 151 AA.
04  PS5226;
05  DT 01-OCT-1996 (Rel. 34, Created)
06  DT 01-OCT-1996 (Rel. 34, Last sequence update)
07  DT 01-MAR-2002 (Rel. 41, Last annotation update)
08  DE Minor curlin subunit precursor (Fimbriae Ssr17 minor subunit).
09  CS CSGB OR AGFB OR STM1143.
10  OS Salmonella typhimurium, and
11  OS Salmonella enteritidis.
12  CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
13  CC Salmonella.
14  CC NCBI_TaxID=602, 592;
15  RN [1]
16  RP SEQUENCE FROM N.A.
17  RC SPECIES=S.typhimurium; STRAIN=SR-11;
18  RX MEDLINE=96117058; PubMed=9457880;
19  RT Rongling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
20  RT "Curli fibers are highly conserved between Salmonella typhimurium and
21  RT Escherichia coli with respect to operon structure and regulation.";
22  RL J. Bacteriol. 180:722-731(1996).
23  RN [2]
24  RP SEQUENCE FROM N.A.
25  RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
26  RX MEDLINE=21534948; PubMed=11677609;
27  RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
28  RA Courtney L., Portnoiiik S., All J., Dante M., Du F., Hou S., Layman D.,
29  RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
30  RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
31  RA Waterston R., Wilson R.K.;
32  RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
33  RT LT2.";
34  RL Nature 413:852-856(2001).
35  RN [3]
36  RP SEQUENCE FROM N.A.
37  RC SPECIES=S.enteritidis; STRAIN=27655-3B;
38  RX MEDLINE=96146512; PubMed=8550497;
39  RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
40  RA "Salmonella enteritidis agtBAC operon encoding thin, aggregative
41  RT fimbriae";
42  RT J. Bacteriol. 178:662-667(1996).
43  RL -I- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
44  CC COLLED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
45  CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
46  CC FIBRINOCTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
47  CC CURLIN MONOMERS.
48  CC -I- SIMILARITY: STRONG, TO E.COLI CSGB.
49  CC -----
50  CC This SWISS-PROT entry is copyright. It is produced through a collaboration
51  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
52  CC the European Bioinformatics Institute. There are no restrictions on its
53  CC use by non-profit institutions as long as its content is in no way
54  CC modified and this statement is not removed. Usage by and for commercial
55  CC entities requires a license agreement (see http://www.isb.ch/announce/
56  CC or send an email to license@isb-sib.ch).
57  CC -----
58  CC EMBL: AJ002301; CAA05316.1; -
59  CC DR EMBL: AE008749; AAL20073.1; -
60  CC DR EMBL: U43280; AAC43598.1; -
61  CC DR StyGene; SG10609; CSGB.
62  CC Fimbrin; Signal; Complete proteome.
63  CC SIGNAL 1 21 POTENTIAL.
64  CC FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
65  CC SQ SEQUENCE 151 AA; 16182 MW; CQPC53430EDD5D36ID CRC64;
66  CC
67  CC Query Match 14.4%; Score 111.5; DB 1; Length 151;
68  CC Best local Similarity 30.5%; Pred. No. 0.021;
69  CC Matches 36; Conservative 17; Mismatches 46; Indels 19; Gaps 5
70  CC
71  CC 51 GSANNAALQSDARKSE-----TTTQSGYGNAGVQ-GADNST-----IELTQ 94
72  CC
73  CC
74  CC
75  CC
76  CC
77  CC
78  CC
79  CC
80  CC
81  CC
82  CC
83  CC
84  CC
85  CC
86  CC
87  CC
88  CC
89  CC
90  CC
91  CC
92  CC
93  CC
94  CC
95  CC
96  CC
97  CC
98  CC
99  CC

```

```

Db 14 GAPGATATNTVDLARSEVNFVAVNELSKSFNFQAMIIQGVGTGDN SARVRQEGSKILLSVISQ 73
Oy 95 NGFRNNATITDWMANKNSDIT--VGQYGGNNALVNQTSADSSVWVRQVFGNNATANQY 151
Db 74 EGNRRARAVDQ--AGNYNFAYIEQTGNANDASISQASAGSAIIQKSGKNRANITQY 129

RESULT 5
OMPB_RICUA STANDARD; PRT; 1656 AA.
AC 006653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen 5) (Scas) (rompb)
DE (comp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptide].
CN OMPB.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YH;
RA Uchiyama T.;
RT "Sequencing of the gene encoding the protein romp B of Rickettsia
RT japonica."
RL Submitted (MAY-1997) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC SPURCULAR PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIOLENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC SIMILARITY).
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce,
CC or send an email to license@slb.ch).
CC -----
DR EMBL; AB003681; BAA20138.1; -
DR InterPro: IPR003658; rompa, rompb.
DR Pfam: PF02708; rompa, rompb, 1.
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1338 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1339 1656 32 KDA BETA PEPTIDE.
FT DOMAIN 528 533 POLY-GLY.
SQ SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;

Query Match 13.3%; Score 103; DB 1; Length 1656;
Best Local Similarity 26.2%; Pred. No. 1,2; Mismatches 52; Indels 56; Gaps
Matches 45; Conservative 19; Gaps 547

Oy 66 SETTTTQSG---YNGADVGGAGDNSTIEITONGFRNNATID----- 104
Db 509 VLAAGATITLDSKTI-----TGDIGNGG-----GAALQSTILANDARK 547
Oy 548 ---TLTILGANNIISANGGTINFGANGTIKLTST--QNNIVPDCDLAIAITDQGVVDASS 602
Db 509 VLAAGATITLDSKTI-----TGDIGNGG-----GAALQSTILANDARK 547
Oy 105 OMANKNSDI--TVGQYGGNNAL-----VNQTSADSSVWVRQVFGNNAT 147
Db 603 LTNAGQILLTISGTIGIIGANNTTILGQFNIGSSKTTLLNGVAINIELVYIANGNS 654

```

DR EMBL: L04979; AAA23616.1; -
 DR EMBL: X90754; CAA62282.1; -
 DR EMBL: AE000205; AAC74126.1; -
 DR EMBL: D90741; BAA35832.1; -
 DR EMBL: D90742; BAA35840.1; -
 DR Ecocore; EG11489; csqa.
 Fimbrin; signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
 FT CONFLICT 7 7 A -> E (IN REF. 1)
 SQ SEQUENCE 151 AA: 15049 MW: C003470D208D395F CRC64;

Query Match 75.8%; Score 586; DB 1; Length 151;
 Best Local Similarity 74.8%; Pred. No. 2e-39;
 Matches 113; Conservative 19; Mismatches 19; Indels 0; Gaps 0;

QY 1 MLTKAAATVYSSALAGVPPOMGGNGGNSGSPDSTLITYGSANALALQ 60
 DB 1 MLTKAAATVYSSALAGVPPOMGGNGGNSGSPDSTLITYGSANALALQ 60
 QY 61 SPARKSETTITQSGYNGADVGAGADNSTELTONGFRNNATIDOMANKSDITVGQYG 120
 DB 61 TPARNSDLTITQHGNGAGADVGGSDSSITDITQKFGNSATLDDMNGKNSKTYAKQFSG 120
 QY 121 NNAALVNOTASDSVYRVQVFGNNATANQY 151
 DB 121 GNGAAVDQATASNSVYVTVQVFGNNATANQY 151

RESULT 3

CSGB_ECOLI STANDARD: PRT; 151 AA.
 AC P39828;

DT 01-FEB-1995 (rel. 31, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Minor curli subunit precursor.
 GN CSGB OR B1041 OR Z1675 OR ECS1419.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OC NCBI_TaxID=562, 83334;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MC4100;
 RX MEDLINE=96414468; PubMed=8817489;
 RA Hammar M., Arngqvist A., Bian Z., Olsen A., Normark S.;
 RT "Expression of two csq operons is required for production of
 RT fibronectin- and congo red-binding curli polymers in Escherichia coli
 RT K-12.";
 RL Mol. Microbiol. 18:661-670(1995).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshimo T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiiuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome

RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN (4)
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grodebeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoustis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blatner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN (5)
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN (6)
 RP SEQUENCE OF 1-21 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=95157246; PubMed=7854117;
 RA Arngqvist A., Olsen A., Normark S.;
 RT "Sigma S-dependent growth-phase induction of the csqBA promoter in
 RT Escherichia coli can be achieved in vivo by sigma 70 in the absence
 RT of the nucleoid-associated protein H-NS.";
 RL Mol. Microbiol. 13:1021-1032(1994).
 CC -I- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -I- SIMILARITY: STRONG, TO SALMONELLA CSGB.
 CC -I- SIMILARITY: TO CURLIN MAJOR SUBUNIT (CSGA).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X90754; CAA62281.1; -
 DR EMBL: AE000205; AAC74125.1; -
 DR EMBL: D90741; BAA35831.1; -
 DR EMBL: AE005315; AAG55787.1; -
 DR EMBL: AP002554; BAB34842.1; -
 DR Ecocore; EG12621; csqa.
 KW Fimbrin; signal; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 151 POTENTIAL.
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 SQ SEQUENCE 151 AA: 15882 MW: B18D266B864014B8 CRC64;

Query Match 14.6%; Score 112.5; DB 1; Length 151;
 Best Local Similarity 30.7%; Pred. No. 0.017;
 Matches 35; Conservative 13; Mismatches 61; Indels 5; Gaps 2;

QY 38 SSGPDSTLITYGSAANALALQSDARKSETTITQSGYNGADVGAGADNSTELTONGF 97
 DB 21 AAGVDLANSEVNF-----AVNELSKSSFNQAAIIOAGATGNNAGLRQGGSKLLAVADEGS 76
 QY 98 RNATITDOMNAKNSDITVYGQVGNNAALVNOTASDSVYRVQVFGNNATANQY 151
 DB 77 SNRAKIDDTGTYNL-AYIDQASANDASISOGAGVNTAMIIQKSGKNANITQY 129


```

RT fimbriae."
RL J. Bacteriol. 178:662-667(1996).
RN [5]
RP SEQUENCE OF 21-151 FROM N.A.
RC SPECIES-S. enteritidis: STRAIN-27655-3B;
RX MEDLINE-94013373; PubMed-8104955;
RA Moran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,
RA Munro C.K., Kay C.M., Bansen P.A., Peterkin P.I., Kay W.W.;
RT "DNA-based diagnostic tests for Salmonella species targeting agfa,
RL the structural gene for thin, aggregative fimbriae."
RL J. Clin. Microbiol. 31:2263-2273(1993).
RN [6]
RP SEQUENCE OF 21-33.
RC SPECIES-S. enteritidis: STRAIN-27655-3B;
RX MEDLINE-91310586; PubMed-1677357;
RA Collinson S.K., Emeedy L., Mueller K.M., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
RT Salmonella enteritidis."
RT J. Bacteriol. 173:4773-4781(1991).
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -1- SIMILARITY: STRONG. TO E. COLI CSGA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ002301; CA005317.1; -
DR EMBL: AF008749; AL20074.1; -
DR EMBL: AL627269; CAP08268.1; -
DR EMBL: U43280; AAC43599.1; -
DR StyGene: SG10608; csqa.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFLICT 134 151 SYMROVGFPGNNATANOY -> DSYTQVAS (IN
FT DNA RES. 3:137-155(1996).
SQ SEQUENCE 151 AA: 15305 MW: 87DAC0D16B621359 CRC64;
Query Match 100.0%; Score 773; DB 1; Length 151;
Best Local Similarity 100.0%; Pred. No. 5, 2e-54;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 KKLKVAFAAIVVSGSALAGVPMQGGGNGHNGGSSGPDSTLSIYQGSANALALQ 60
DB 1 KKLKVAFAAIVVSGSALAGVPMQGGGNGHNGGSSGPDSTLSIYQGSANALALQ 60
OY 61 SPARKSETTITQSGCGNADVGQGGDNSTITELTONGCFRNNTTIOMNKNSDITVGOYGG 120
DB 61 SPARKSETTITQSGCGNADVGQGGDNSTITELTONGCFRNNTTIOMNKNSDITVGOYGG 120
OY 121 NNAALVNOTASSDSYVNRQVGFSGNNATANOY 151
DB 121 NNAALVNOTASSDSYVNRQVGFSGNNATANOY 151
RESULT 2
CSGA_ECOLI STANDARD: PRT: 151 AA.
AC 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR B1042.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

```

```

OC Escherichia
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / W3110;
RX MEDLINE-93211294; PubMed-8459772;
RA Olsen A., Arngvist A.;
RT "The RpoS sigma factor relieves H-NS-mediated transcriptional
RT repression of csqa, the subunit gene of fibronectin-binding curli in
RT Escherichia coli."
RL Mol. Microbiol. 7:523-536(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MC4100;
RX MEDLINE-9641468; PubMed-8817489;
RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csq operons is required for production of
RT fibronectin- and congo red-binding curli polymers in Escherichia coli
RT K-12."
RL Mol. Microbiol. 18:661-670(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-97061202; PubMed-8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Iemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
RN [5]
RP SEQUENCE OF 21-40.
RC STRAIN-K12 / YMBL;
RX MEDLINE-93023873; PubMed-1357528;
RA Arngvist A., Olsen A., Pfeiffer J., Russell D.G., Normark S.;
RT "The Crl protein activates cryptic genes for curli formation and
RT fibronectin binding in Escherichia coli HB101."
RL Mol. Microbiol. 6:2443-2452(1992).
RN [6]
RP SEQUENCE OF 21-31.
RC MEDLINE-91310586; PubMed-1677357;
RA Collinson S.K., Emeedy L., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
RT Salmonella enteritidis."
RL J. Bacteriol. 173:4773-4781(1991).
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -1- SIMILARITY: STRONG. TO SALMONELLA CSGA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 11, 2002, 21:05:23 ; Search time 10 seconds
(without alignments)
584.665 Million cell updates/sec

Title: US-09-543-407-5
Perfect score: 773
Sequence: 1 MKLLKVAAPFAIVSSGALA.....DSSVMVRQVCGNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	773	100.0	151	1	CSGA_SALTY
2	586	75.8	151	1	P55225 salmonella
3	112.5	14.6	151	1	P28307 escherichia
4	111.5	14.4	151	1	P39828 escherichia
5	103	13.3	1656	1	P55226 salmonella
6	101.5	13.1	1567	1	O06653 r outer mem
7	98.5	12.7	1028	1	P18127 xanthomonas
8	98	12.7	678	1	P51521 drosophila
9	96	12.4	959	1	Q10778 mycobacteri
10	95.5	12.4	1034	1	Q02629 saccharomyc
11	95.5	12.4	1258	1	Q47879 pantoea ana
12	93.5	12.1	1322	1	P16239 erwinia her
13	92.5	12.0	1317	1	P20469 pantoea ana
14	92	11.9	1317	1	P49687 saccharomyc
15	89.5	11.6	1655	1	O9Kka3 r outer mem
16	89.5	11.6	262	1	P07875 bacterioph
17	89.5	11.6	760	1	P75780 escherichia
18	89.5	11.6	1196	1	ICEV_PSEEX
19	89	11.5	401	1	P34291 pseudomonas
20	88.5	11.4	1210	1	P08815 caenorhabdi
21	87.5	11.3	491	1	P08815 pseudomonas
22	87.5	11.3	487	1	Q10778 mycobacteri
23	87.5	11.3	1148	1	P47611 mycobacteri
24	87.5	11.3	1200	1	O36011 pseudomonas
25	87	11.3	823	1	P06620 pseudomonas
26	86	11.1	347	1	P1907 saccharomyc
27	86	11.1	1093	1	Q03646 plasmidium
28	85.5	11.1	641	1	Q04052 drosophila
29	85.5	11.0	796	1	Q44052 drosophila
30	84.5	10.9	1140	1	Q04891 candida alb
31	84	10.9	485	1	O81172 treponema p
32	83.5	10.8	894	1	Q12906 h interleaf
33	83.5	10.8	1067	1	SGG_DROME

34	83.5	10.8	1113	1	N116_YEAST	Q02630 saccharomyc
35	83.5	10.8	1156	1	GLH4_CAEEL	O76743 caenorhabdi
36	83	10.7	342	1	OMP_C_RAHQ	O33507 rahnella ag
37	83	10.7	429	1	DR48_YEAST	P18899 saccharomyc
38	83	10.7	543	1	YP91_MYCTU	Q0630 mycobacteri
39	82.5	10.7	369	1	PST3_MYCAV	O9Kk89 mycobacteri
40	81.5	10.5	392	1	HME1_HUMAN	O05925 homo sapien
41	81.5	10.5	548	1	CEAK_ECOCI	Q47502 escherichia
42	81.5	10.5	1063	1	SPT5_YEAST	P27697 saccharomyc
43	81	10.5	165	1	GRP1_ORISA	P25074 oryza sativ
44	81	10.5	400	1	RTOA_DICDI	P54681 dictyostell
45	81	10.5	443	1	Y878_MYCTU	Q10540 mycobacteri

ALIGNMENTS

RESULT 1	ID	CSGA_SALTY	STANDARD	PRT	151 AA.
AC	P55225				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	01-MAR-2002 (Rel. 41, Last annotation update)				
DE	Major curlin subunit precursor (Fimbrin SEF17).				
GN	CSGA OR AGEA OR STM1144 OR STY1181.				
OS	Salmonella typhimurium,				
OS	Salmonella typhi, and				
OS	Salmonella enteritidis.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC	Salmonella.				
OX	NCBI_Taxid=602, 601, 592;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S.typhimurium; STRAIN=SR-11;				
RX	MEDLINE=98117058; PubMed=9457880;				
RA	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;				
RT	"Curli fibers are highly conserved between Salmonella typhimurium and				
RT	Escherichia coli with respect to operon structure and regulation.";				
RL	J. Bacteriol. 180:722-731(1998).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;				
RX	MEDLINE=21534948; PubMed=11677609;				
RA	McClelland M., Sanderon K.E., Spleth J., Clifton S.W., Latelle P.,				
RA	Courtney L., Potwolik S., Ali J., Dante M., Du F., Hou S., Layman D.,				
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Molyaney E.,				
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,				
RA	Waterston R., Wilson R.K.;				
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium				
RT	LT2.";				
RL	Nature 413:852-856(2001).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S.typhi; STRAIN=CT18;				
RX	MEDLINE=21534947; PubMed=11677608;				
RA	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,				
RA	Churcher S., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,				
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,				
RA	Ciclin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,				
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jajels K.,				
RA	Krog A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,				
RA	Quill M., Rutherford K., Simmonds M., Skelton J., Stevens K.,				
RA	Whitehead S., Barrett B.G.;				
RT	"Complete genome sequence of a multiple drug resistant Salmonella				
RT	enterica serovar Typhi CT18.";				
RL	Nature 413:848-852(2001).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S enteritidis; STRAIN=27655-3B;				
RX	MEDLINE=96146512; PubMed=8550497;				
RA	Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;				
RT	"Salmonella enteritidis agfBac operon encoding thin, aggregative				